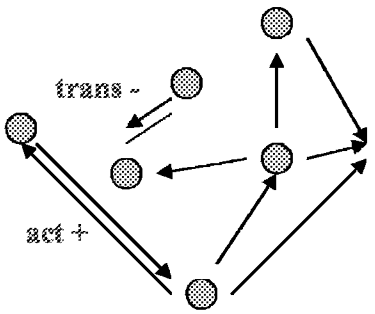
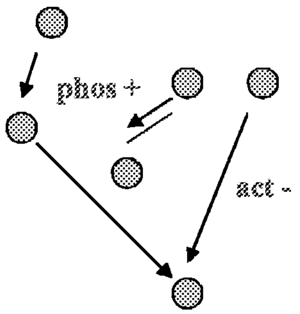
EXHIBIT 1

Analyzing Expression Results

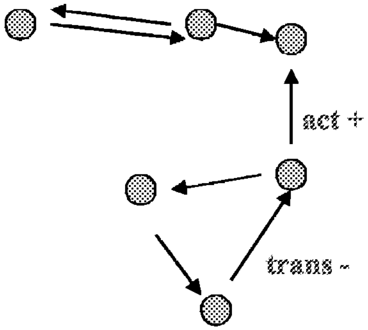
An underlying belief that the cell works in pathways….



cell death



wound-healing

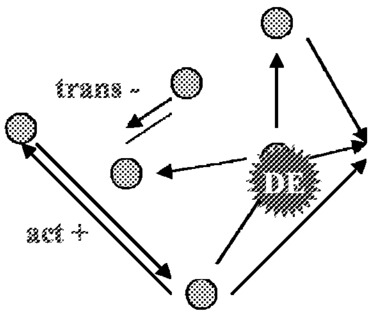


glucose metabolism

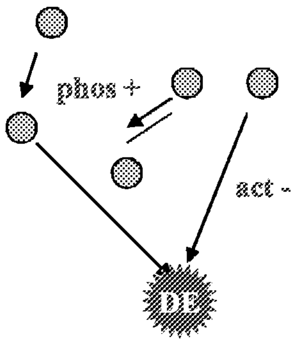
EXHIBIT 1 Con't.

Analyzing Expression Results

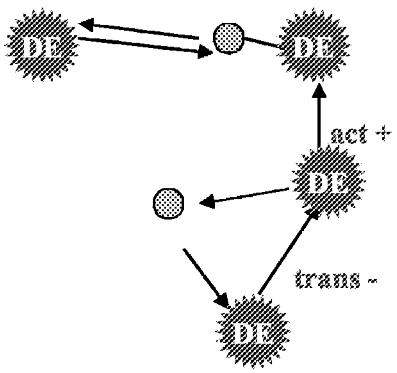
and the use of expression data to identify which pathways are disrupted….



cell death



wound-healing



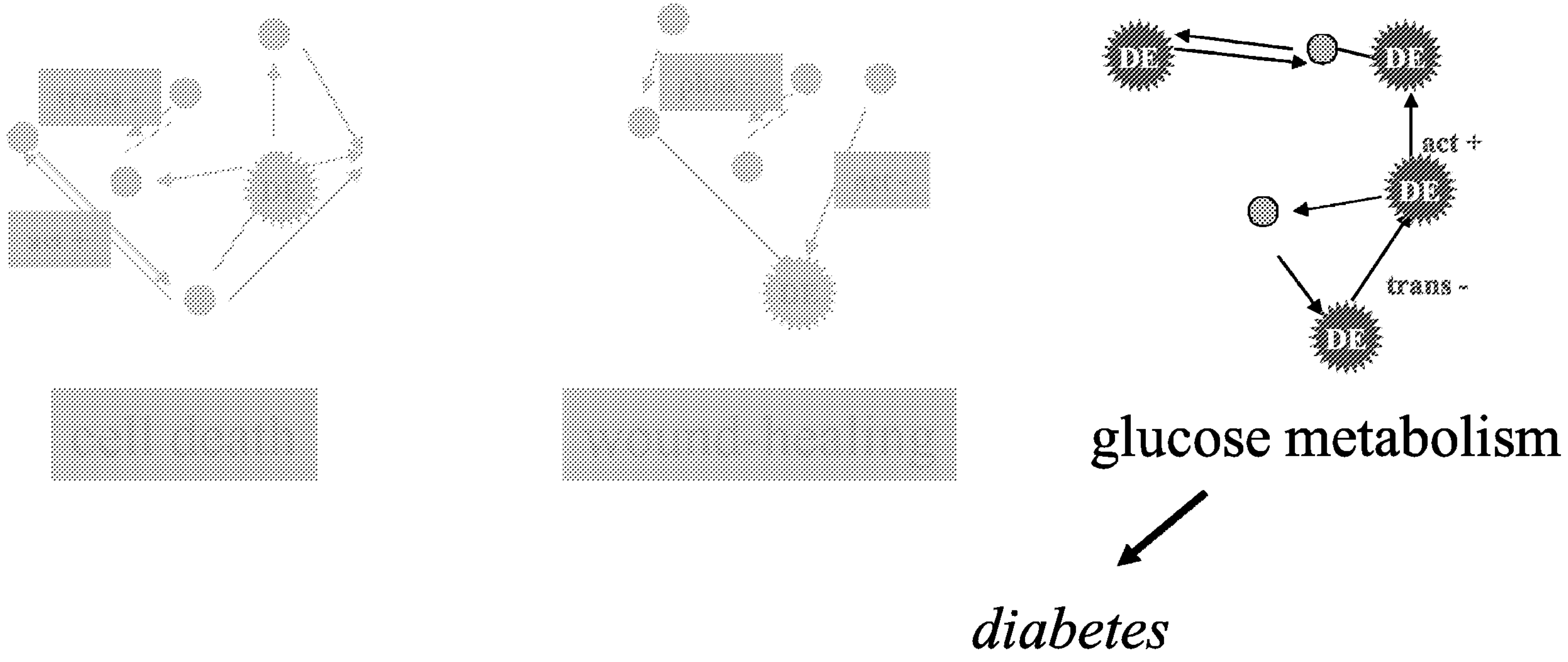
glucose metabolism

expression profile of diabetes

EXHIBIT 1 Con't.

Analyzing Expression Results

in a disease state so that small molecule therapeutics can be prioritized and tested.



glucose metabolism

diabetes

EXHIBIT 1 Con't.

Analyzing Expression Results

Q: What need do our customers need addressed for them?

A: Connections between expression results and biological pathways that do not appear to be the product of random chance.

EXHIBIT 1 Con't.

Analyzing Expression Results

Q: Why are connections between expression results and biological pathways valuable?

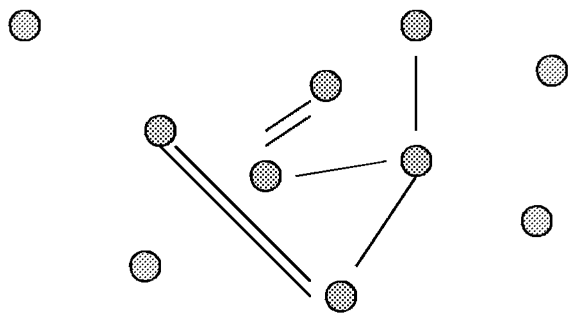
A: A pathway specifically disrupted in a disease offers a set of genes that can be targeted in a specific way by therapeutic discovery scientists to help ameliorate the effects of the disease.

EXHIBIT 1 Con't.

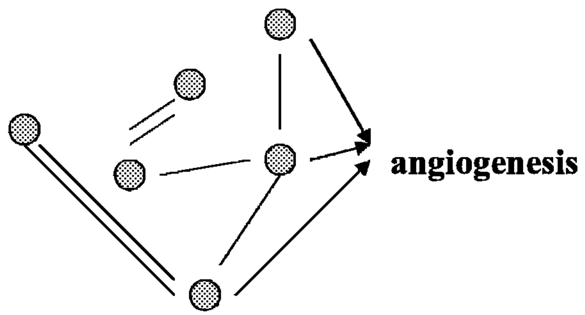
Q: What are different ways biological pathways can be defined?

A:

1. Sets of genes that are more functionally interactive with one another, quantitatively and qualitatively, than other genes in the genome…



2. …that are linked to discrete cellular activities…



3. …in a mechanistically compelling way.

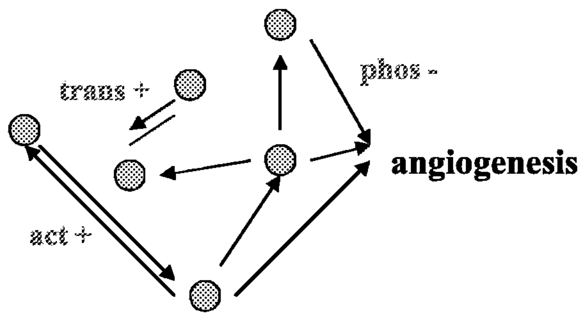
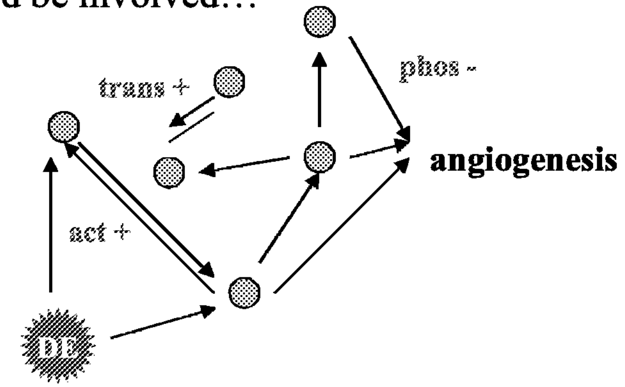


EXHIBIT 1 Con't.

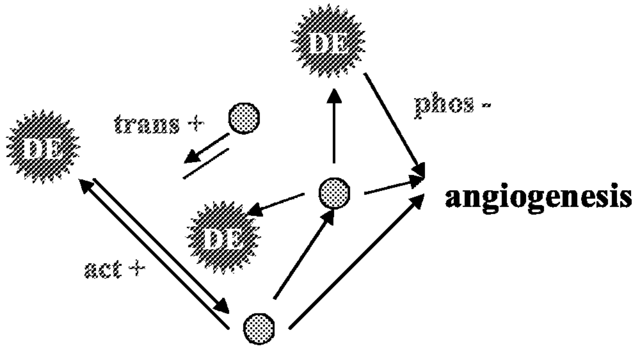
Q: What are different ways expression data could be connected to pathways?

A:

2. …but as our computational sophistication grows, inference could be involved…



1. Most simply, appearance of Differentially Expressed genes in pathways…



3. …and also mechanism of action.

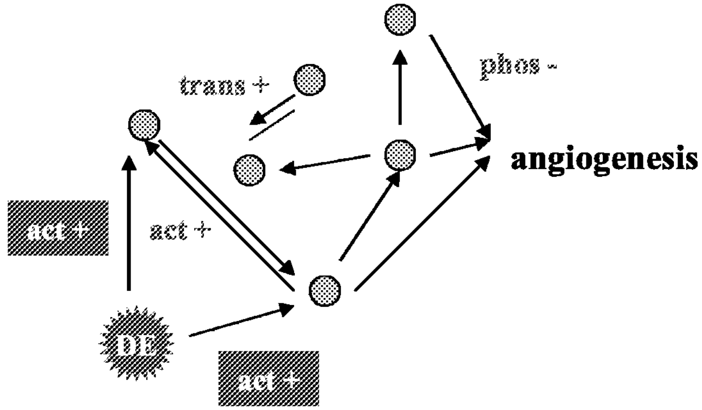
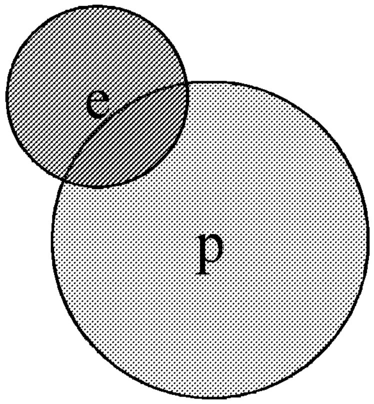


EXHIBIT 1 Con't.

Q: What does it mean that a connection is non-random?

A: That there is some evidence that the connection between the expression data set and the pathway is not just a chance occurrence we are making hoopla over to make a sale.



There are probably more than a hundred ways we could make such a case; a fairly simplistic set of equations asks the following question: *What are the chances that the observed overlap between members of set* **e** *(expression data) and* **p** *(pathway) is due to random chance?*

EXHIBIT 1 Con't.

Goal of EAFX Project

Establish an algorithm or set of algorithms that:

-Defines pathways in a way that is customer-validated

-Identifies connections between expression results and these pathways

-Is able to identify those connections that do not appear to be the product of random chance

EXHIBIT 1 Con't.

Sample Proposal

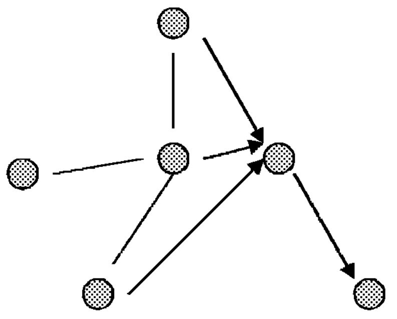
(Expression data input for small molecule or disease profile)

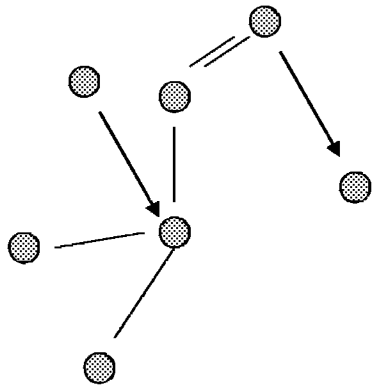
Gene	Fold diff	Direction
A	2.4	+
B	3.6	+
C	1.2	-
D	0.2	-
E	5.7	+
F	0.2	-
G	8.2	+

EXHIBIT 1 Con't.

Sample Proposal

(Possible output)





Total absolute expression delta:	20.3	17.1
Total findings:	32	41
Total edges:	7	6
Process overrep:	Immune activation	Apoptosis
	(1 x 10^{-4})	(3 x 10^{-8})
Overall score:	45	43
Overall freq. of this score or higher:	1/1000	1/230
Cluster rank:	1	2

EXHIBIT 1 Con't.

Key questions to be asked of EAFX output, in order of priority

1. What gene-based pathways are disrupted in this disease (or by this small molecule)?
2. How is this gene-based pathway disrupted in this disease?
3. What are central processes affected by the genes in this pathway?
4. How are these processes affected by the genes in this pathway?
5. What are small molecules that exert an effect on genes in this pathway?
6. How can I test whether a particular small molecule is restoring “normal function” of this pathway on a genetic level?
7. What genes in this pathway have been patented (and for what purpose)?
8. What genes in this pathway have a link to cellular or organismal toxicity?

EXHIBIT 2

EAFX Proposal
Draft:

1) **Reiterate EAFX Project Goal:**
Use Ingenuity's content to identify connections between expression results and biological pathways that do not appear to be the product of random chance.

2) Suggest range of possible solutions.
From simple to futuristic.

Difficulty	**Solution**
Basic (Implementable)	<u>Simple relationships between genes</u> Take set of genes-> Identify all direct facts linking genes. Identify largest connected groupings. Identify links with lots of facts.
Realistic but risker than supervised approach.	Unsupervised approach to identifying clusters. Develop algorithms that can automatically identify functional clusters based on correlations between user genes and knowledge connectedness/densities in our kb.
Futuristic (Science-fiction)	Create a virtual model of tissue/disease specific cells using expanded Ingenuity structured content (scientific literature, genomics data, bioinformatics data, canonical knowledge, user knowledge, pre-existing analysis). Develop sophisticated algorithms that predict behavior and that identify mechanistic explanations for dysregulated pathways.

EXHIBIT 2 Con't.

3) Define improvement axis (biologically believable, significance likelihood, decision relevance)

The value of our product increases by improving the user's x,y or z with the "pathways" generated by our analysis.

1) Biologically believable: The results are consistent with the user's understanding of biology. (ie. Fill in an example)

2) Significance likelihood: The results do not appear to be the product of random chance. (unique, unexpected, specificity to their input, correlated with input) (i.e. Most of the genes upregulated by a specific transcription factor are among the input genes.)

3) Decision Relevance: The results are applicable to the user's decision-making process.

i.e. Interesting drug discovery traits:

a. Uniqueness/Novelty
b. Patent
c. Tissue Specificity (Link to body atlas)
d. Toxicity
e. Disease

4) Possible features that would "improve" performance/functionality.

INPUT
CONTENT
ALGORITHMS/SCORING

EXHIBIT 2 Con't.

Improvement	Biologically Believable	Significance Likelihood	Decision Relevance	Other Notes
INPUT				
List of Genes				
Cluster	Baseline	Baseline	Baseline	
All measured	~	+++	++	
Cluster membership	+	+++	+++	Assumes belief that clusters have significance
Expression values				
Dir of Change	+++	++	~	
Quantity (1 exp)	++	++	~	
Quantity over time (Time Series)	++++	+++	~	
Experimental Context				
Disregulated Genes	+++	++	++	Knockouts, overexpression
Cell/Tissue Source	+++	++	++	Includes expression specificity
Cell/Tissue disease state				
Cell/Tissue Treatment (Small molecule, irradation)				
CONTENT				
Kb Objects				
Unspecified	Baseline	Baseline	Baseline	
Mutant vs Wildtype				
Localization				
Active vs inactive state				
Complex vs unbound				
Species specificity				

EXHIBIT 2 Con't.

KB Processes
Molecular
Modification
Complex formation

Confidence of link
Negated information
Coupling (indirect vs indirect)

Fact type
Structure
Disease correlation

Include simple example/output that this would allow

5) Internal recommendation:
Define baseline proposal (in addition helps us better understand the system)

Realistic Requires	Supervised approach to identifying clusters. Use expert/algorithmic rules to generate potentially meaningful biological profiles. Scan user's genes against all profiles to identify interesting mechanisms. Refine profiles based on user's particular genes.
Realistic but risker than supervised approach.	Unsupervised approach to identifying clusters. Develop algorithms that can automatically identify functional clusters based on correlations between user genes and knowledge connectedness/densities in our kb.

EXHIBIT 3

I worked out the probability (not p-value) calculation for the null hypothesis match. It is most significantly impacted by the overlap (the number of 'significant' user genes and the KB genes in a particular BCP). I implemented the machine precision-optimized calculation in PERL and checked it into the eafx/scripts directory 'random_match_prob.pl'. Please read below (also in the source file) for details.

Dan

```
# Compute probabily of getting BCP match by chance for null hypothesis
# of BCP generated randomly.
#
# Dan Richards
# [DATE REDACTED]
#
# Inputs:
#   SIG     - number of significant user genes that are mapped to KB genes
#   OVP     - number of (significant) user genes that overlap the KB genes
#             in the BCP
#   MAP     - total number of user genes assayed (not necessarily
significant)
#             that are mapped to KB genes.
#   KB      - number of KB genes (which could appear in a BCP--ie. have
#             suitable content)
#   BCP     - number of KB genes in the BCP
#
# Formula for significance:
#
# 1. P(USER_OVP) = probability that the particular number of overlapping
#                  genes occur in the user's data set
#                = Choose(SIG,OVP) / Choose (MAP,OVP)
# 2. P(BCP_OVP)  = probability that the particular number of overlapping
#                  genes occur in the BCP
#                = (Choose(OVP,OVP) * Choose(KB-OVP,BCP-OVP)) /
Choose(KB,BCP)
#                = Choose(KB-OVP,BCP-OVP) / Choose(KB,BCP)
#                  Note: Choose(OVP,OVP) = 1
# 3. P(OVP)      = P(USER_OVP) and P(BCP_OVP)
#                = P(USER_OVP) * P(BCP_OVP)
#                = (Choose(SIG,OVP) * Choose(KB-OVP,BCP-OVP)) /
#                  (Choose(MAP,OVP) * Choose(KB,BCP))
#
# Implications:
# 1. For a fixed set KB genes, and a fixed number of SIGnificant user genes:
#   a. The larger the BCP, the MORE likely the match occurred by chance
#   b. The larger the OVP, the LESS likely the match occurred by chance
# 2. For a fixed number of OVP genes, and a fixed size of the matched BCP:
#   a. The larger the SIG, the MORE likely the match occurred by chance
#   b. The larger the KB, the LESS likely the match occurred by chance
# 3. If BCP=KB, then if there is any overlap, P(OVP) is unity (1).
# 4. If SIG=KB, then if there is any overlap, P(OVP) is unity, since this
#    implies that every gene in the KB is also significant user gene, so
#    every match is expected.
# 5. If MAP<KB, then the P(OVP) is greater (MORE likely) than if MAP=KB
#
# So overall, P(OVP) is minimized (LEAST likely) if (in decreasing
likelihood):
# KB >> BCP, OVP >> 1, MAP=KB, BCP=OVP, SIG=OVP
#
```

EXHIBIT 3 Con't.

```
# Note: an overlap of more than 1 to a BCP with more than 1 gene is MUCH
#       less probable than an overlap of 1 to a BCP with only 1 gene.
#
# Invariants:
# KB >= 0
# BCP <= KB
# MAP <= KB
# SIG <= MAP
# OVP <= BCP, OVP <= SIG

sub p_overlap {
   # Computes p(OVP) result to highest possible machine precision:
   #
   # P(OVP) formula simplifies to:
   # (SIG! * BCP! * (KB-OVP)! * (MAP-OVP)!) /
   # (SIG-OVP)! * (BCP-OVP)! * KB! * MAP!
   #
   # Note:
   # n! = GAMMA(n+1)
   #
   # Uses log() to maintain highest possible numerical machine precision
   #
   # Non-optimized (equivalent) formula:
   # return (choose($sig,$ovp)*choose($kb-$ovp,$bcp-$ovp))/(choose($map,$ovp)*choose($kb,$bcp));
```

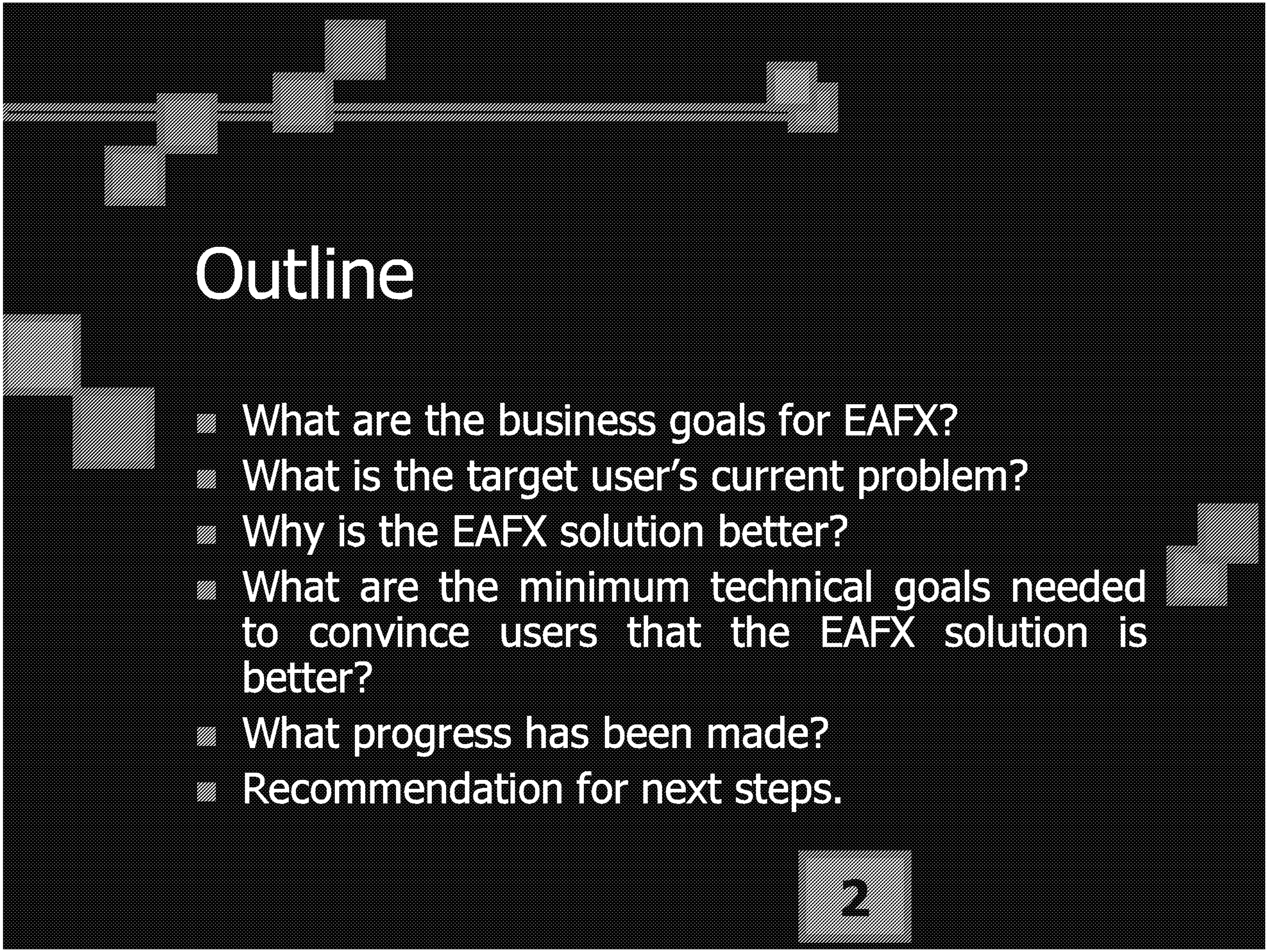
EXHIBIT 4

EAFX Progress Report

David Lin, Ray Cho, Dan Richards, Keith Steward

1

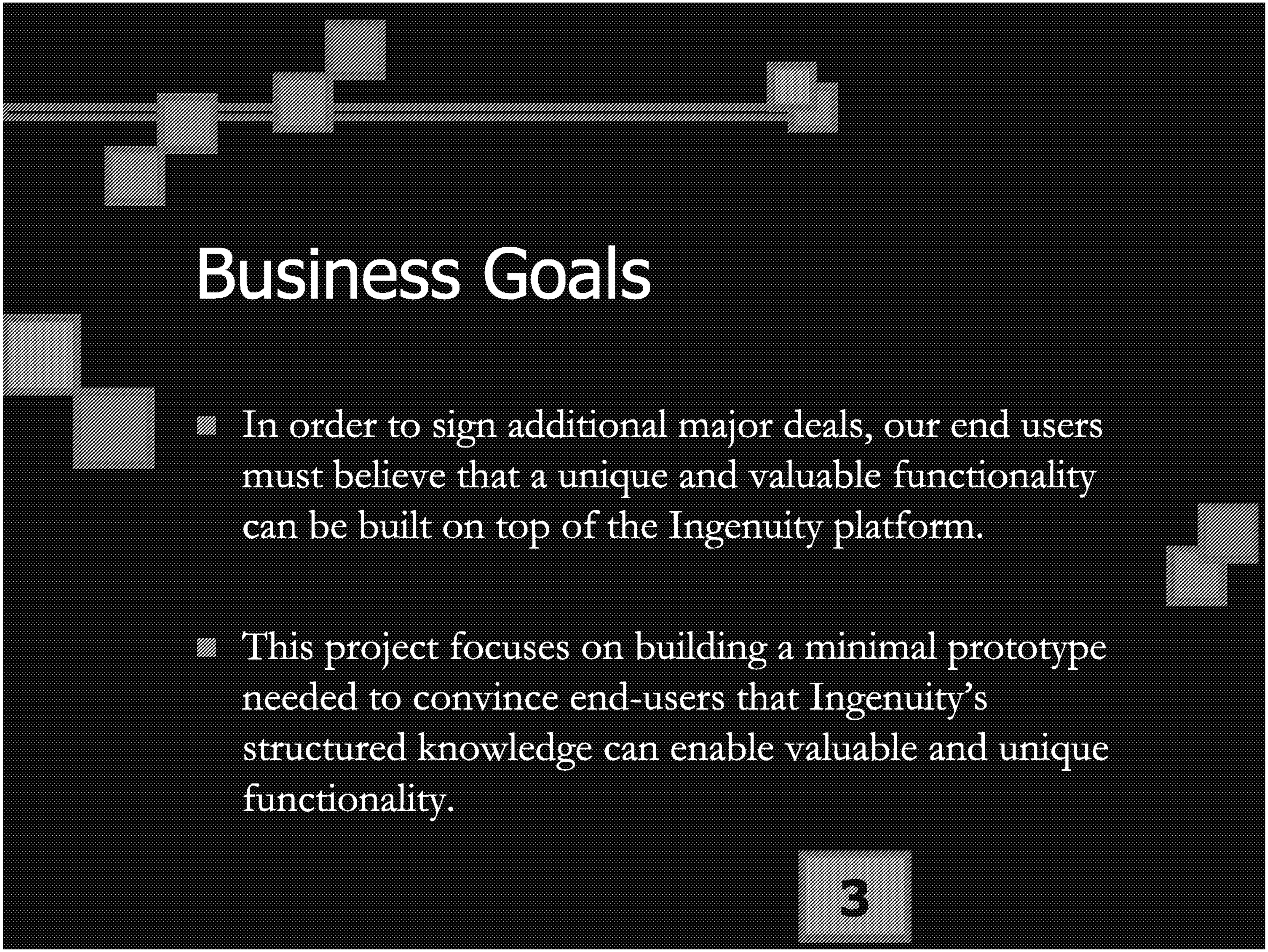
EXHIBIT 4 Con't



Outline

- What are the business goals for EAFX?
- What is the target user’s current problem?
- Why is the EAFX solution better?
- What are the minimum technical goals needed to convince users that the EAFX solution is better?
- What progress has been made?
- Recommendation for next steps.

EXHIBIT 4 Con't

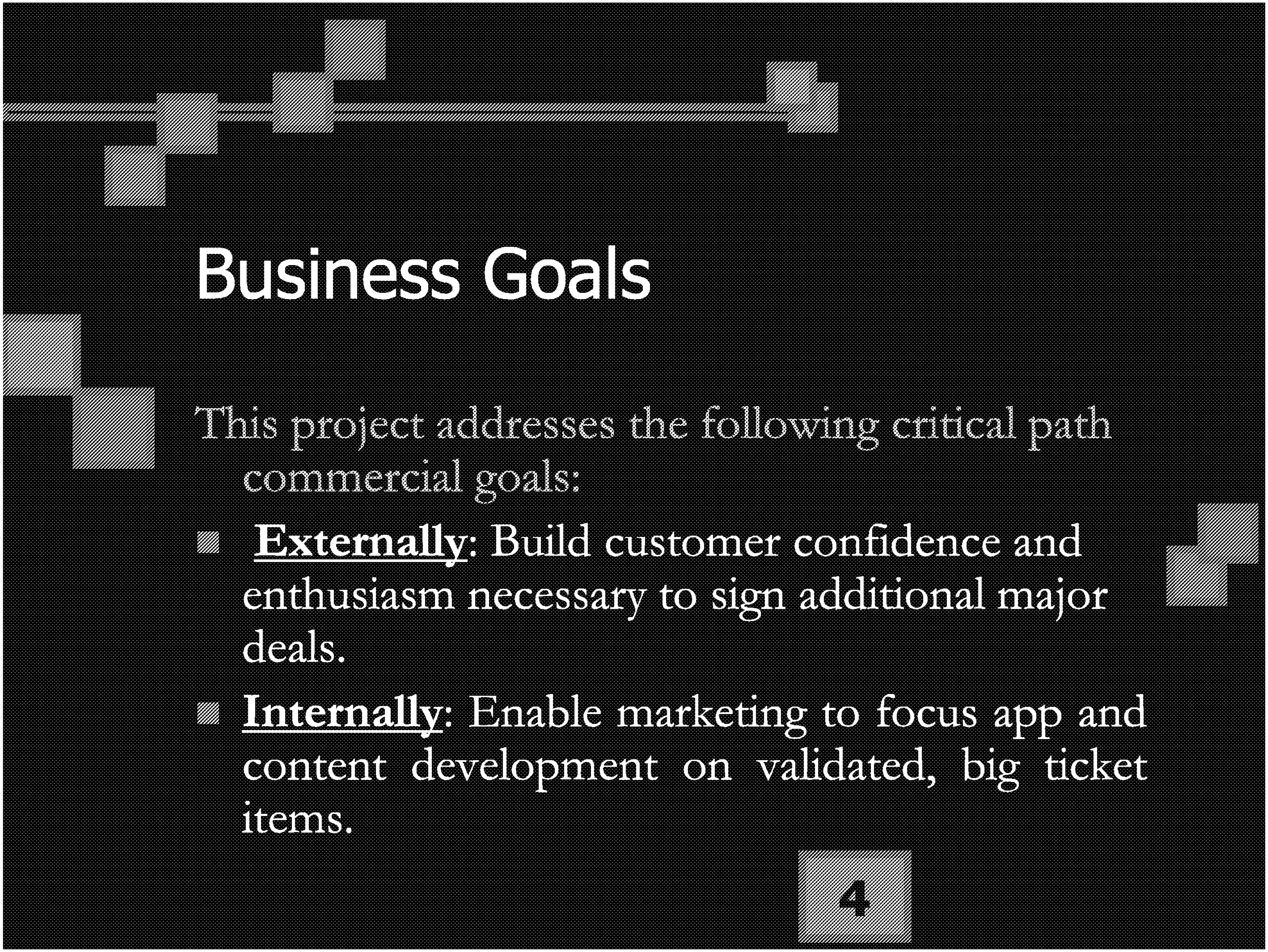


Business Goals

- In order to sign additional major deals, our end users must believe that a unique and valuable functionality can be built on top of the Ingenuity platform.

- This project focuses on building a minimal prototype needed to convince end-users that Ingenuity's structured knowledge can enable valuable and unique functionality.

EXHIBIT 4 Con't

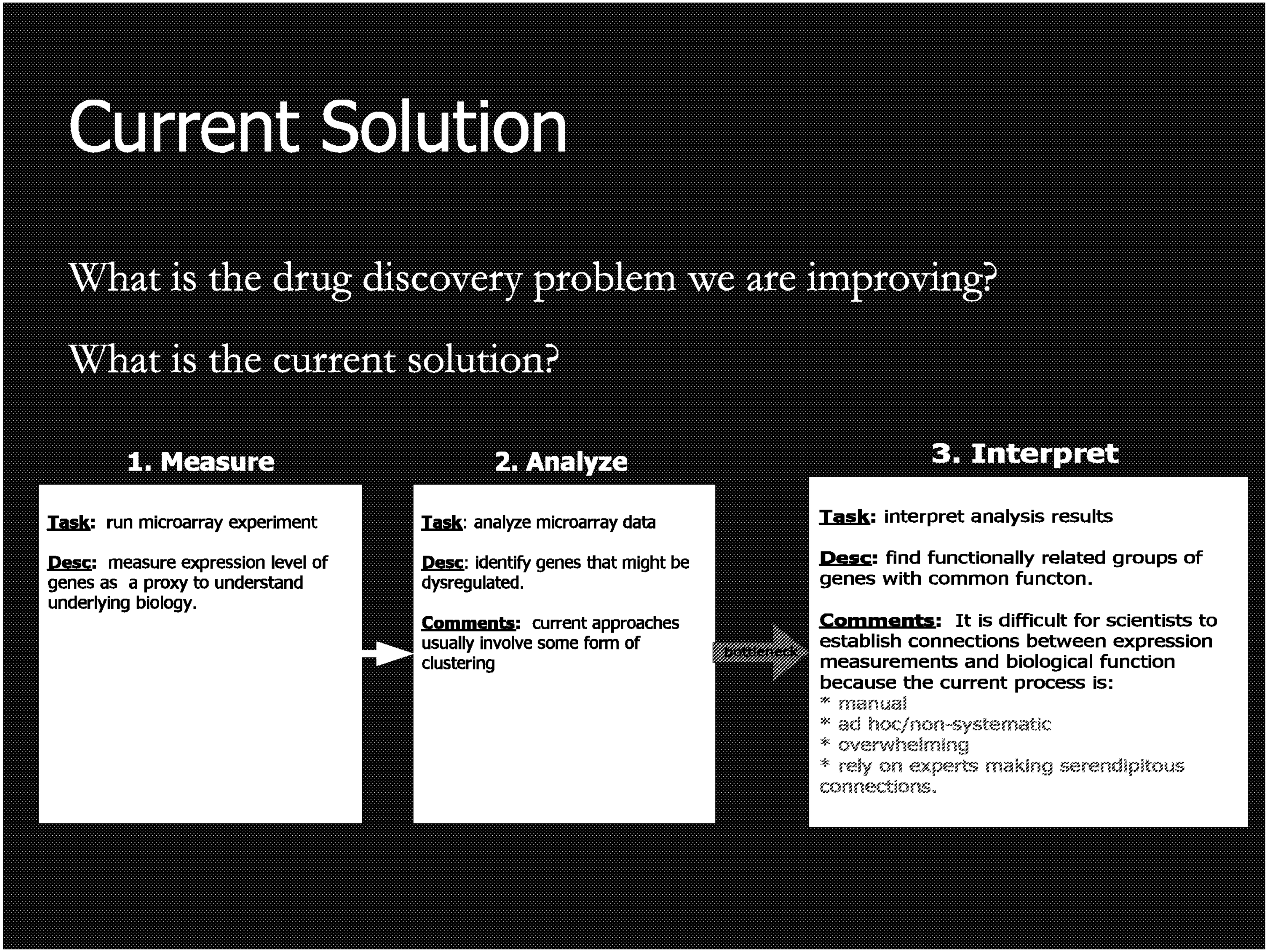


Business Goals

This project addresses the following critical path commercial goals:

- **Externally**: Build customer confidence and enthusiasm necessary to sign additional major deals.
- **Internally**: Enable marketing to focus app and content development on validated, big ticket items.

EXHIBIT 4 Con't



Current Solution

What is the drug discovery problem we are improving?

What is the current solution?

1. Measure

Task: run microarray experiment

Desc: measure expression level of genes as a proxy to understand underlying biology.

2. Analyze

Task: analyze microarray data

Desc: identify genes that might be dysregulated.

Comments: current approaches usually involve some form of clustering

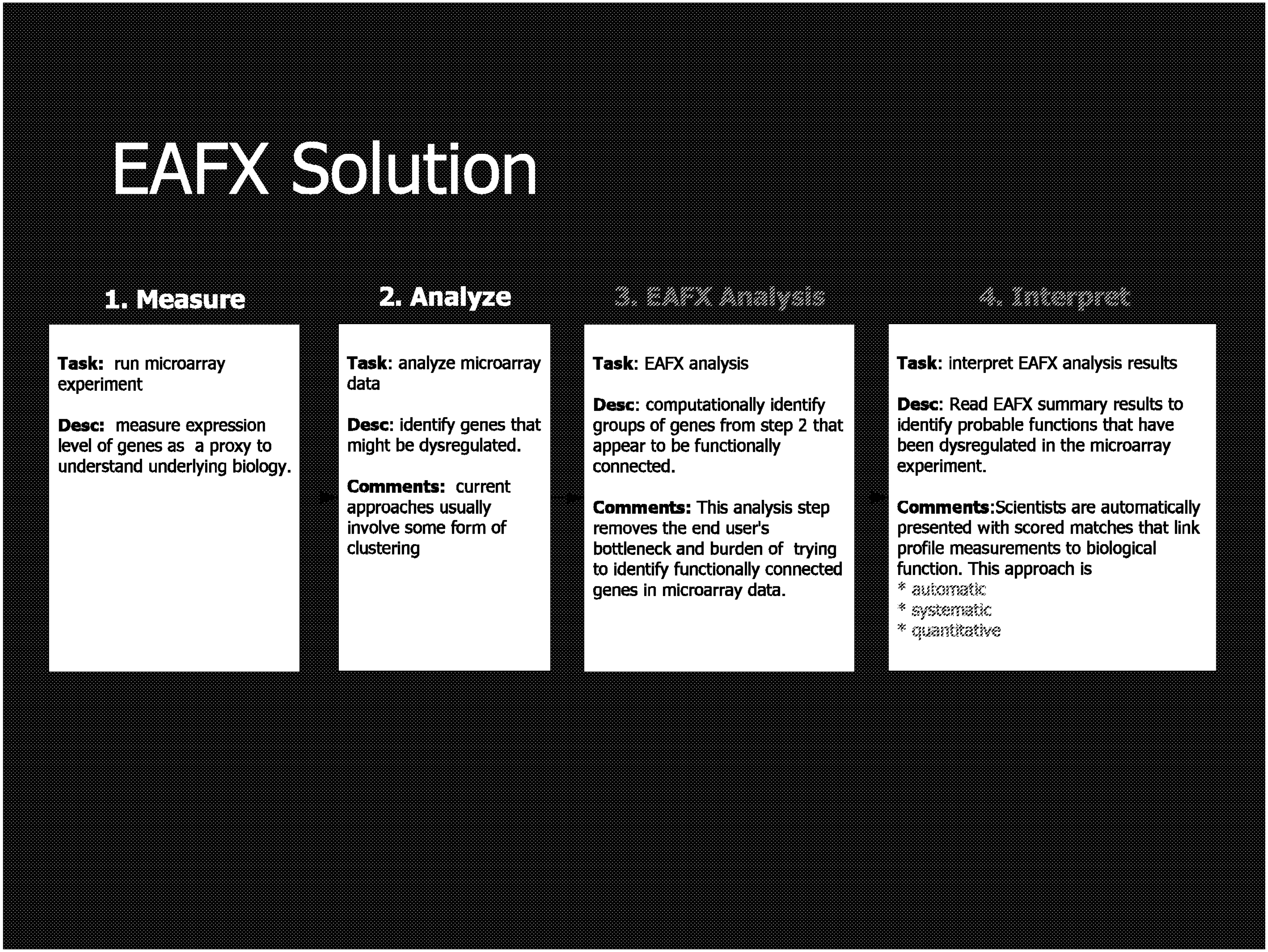
3. Interpret

Task: interpret analysis results

Desc: find functionally related groups of genes with common functon.

Comments: It is difficult for scientists to establish connections between expression measurements and biological function because the current process is:
* manual
* ad hoc/non-systematic
* overwhelming
* rely on experts making serendipitous connections.

EXHIBIT 4 Con't



EAFX Solution

1. Measure

Task: run microarray experiment

Desc: measure expression level of genes as a proxy to understand underlying biology.

2. Analyze

Task: analyze microarray data

Desc: identify genes that might be dysregulated.

Comments: current approaches usually involve some form of clustering

3. EAFX Analysis

Task: EAFX analysis

Desc: computationally identify groups of genes from step 2 that appear to be functionally connected.

Comments: This analysis step removes the end user's bottleneck and burden of trying to identify functionally connected genes in microarray data.

4. Interpret

Task: interpret EAFX analysis results

Desc: Read EAFX summary results to identify probable functions that have been dysregulated in the microarray experiment.

Comments:Scientists are automatically presented with scored matches that link profile measurements to biological function. This approach is
* automatic
* systematic
* quantitative

EXHIBIT 4 Con't

Basis of EAFX

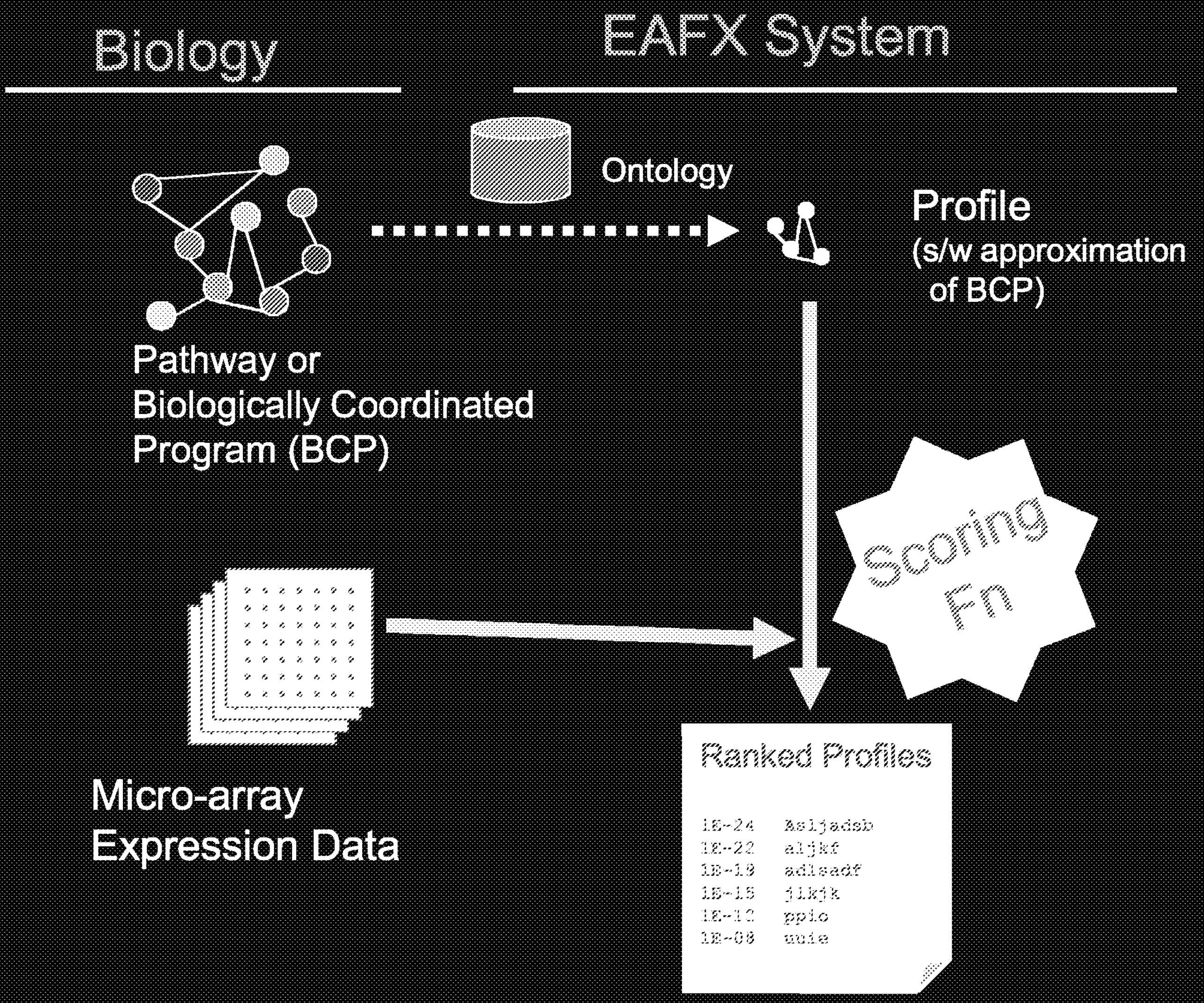
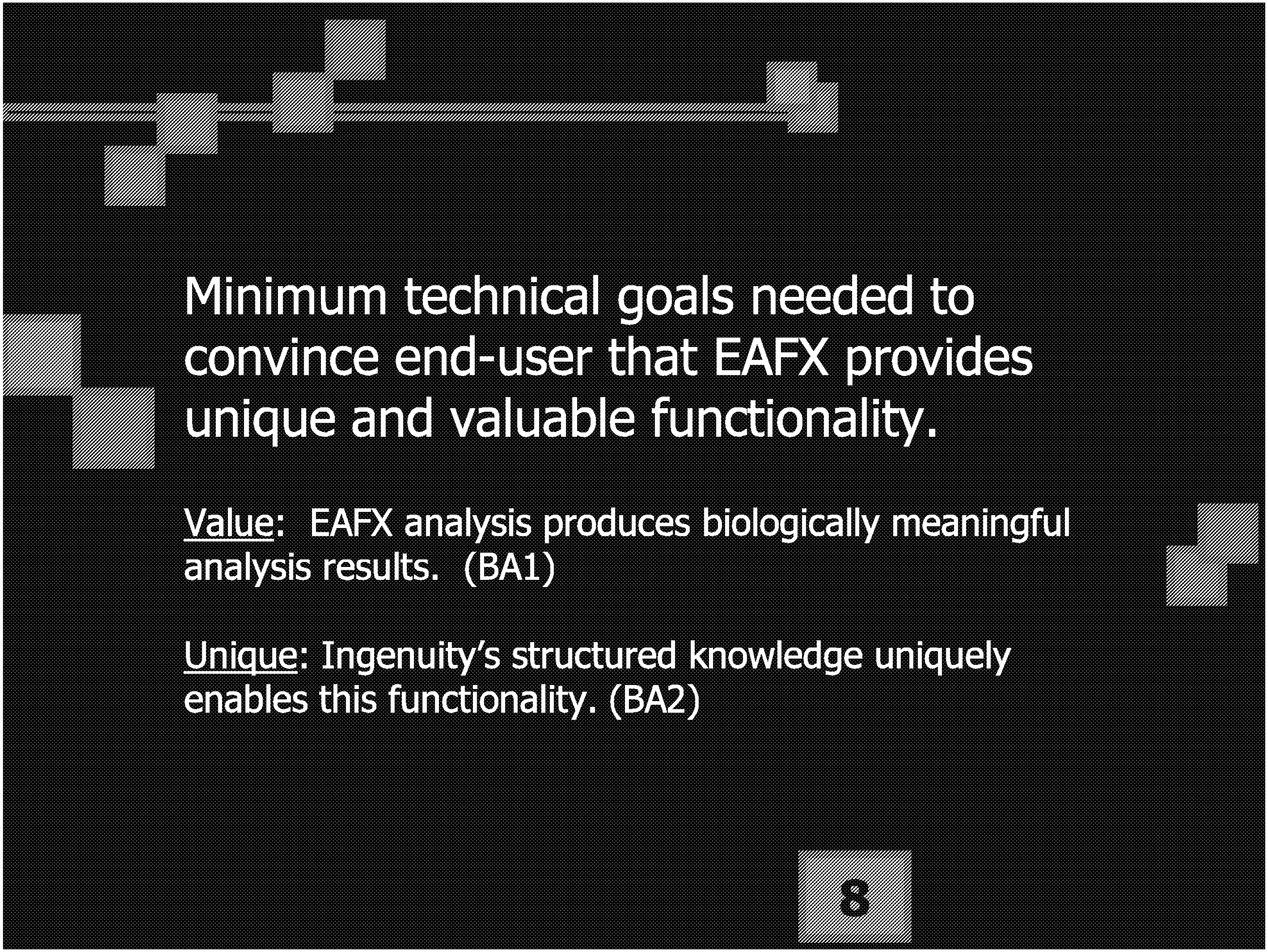


EXHIBIT 4 Con't

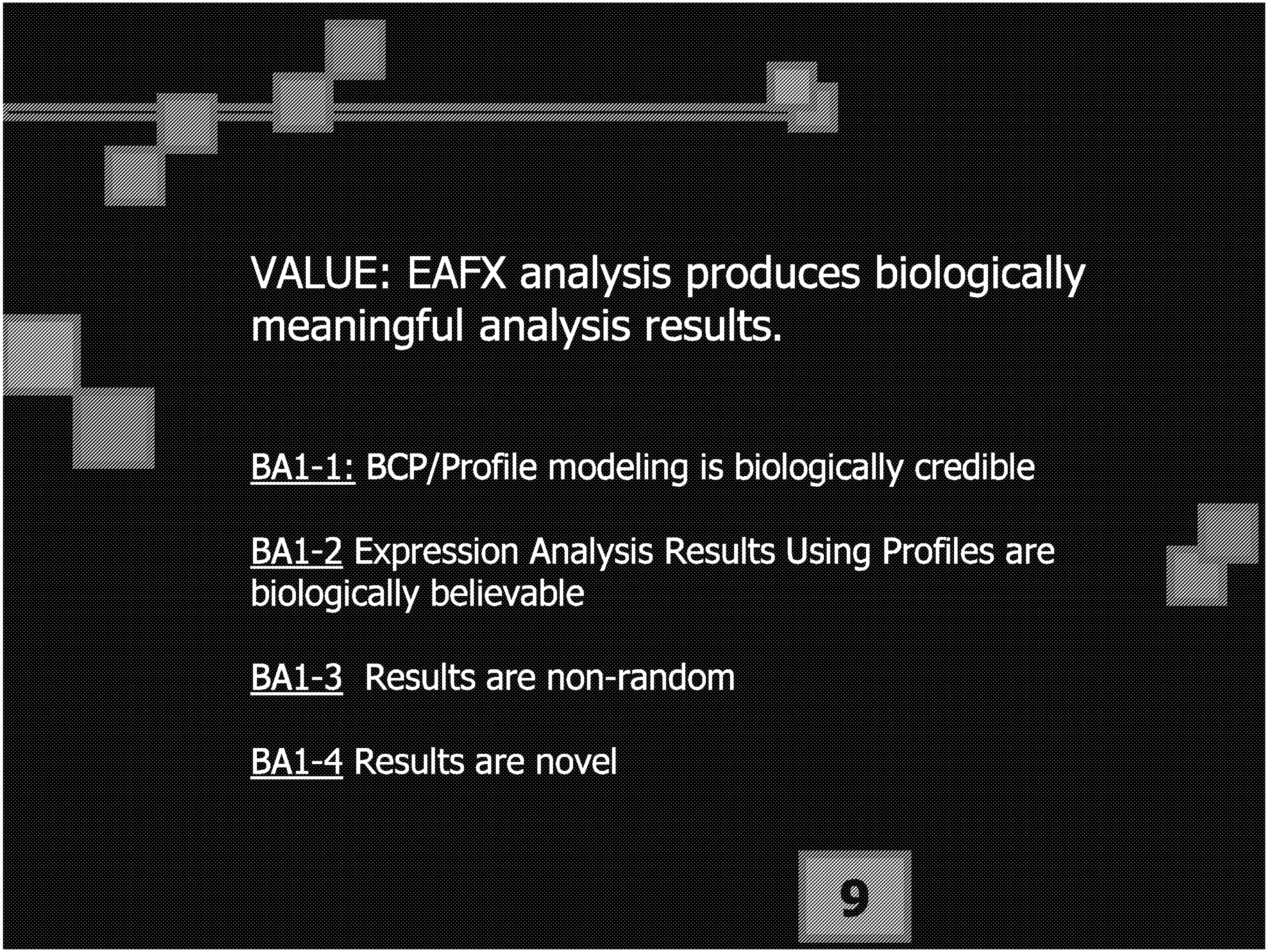


Minimum technical goals needed to convince end-user that EAFX provides unique and valuable functionality.

Value: EAFX analysis produces biologically meaningful analysis results. (BA1)

Unique: Ingenuity's structured knowledge uniquely enables this functionality. (BA2)

EXHIBIT 4 Con't



VALUE: EAFX analysis produces biologically meaningful analysis results.

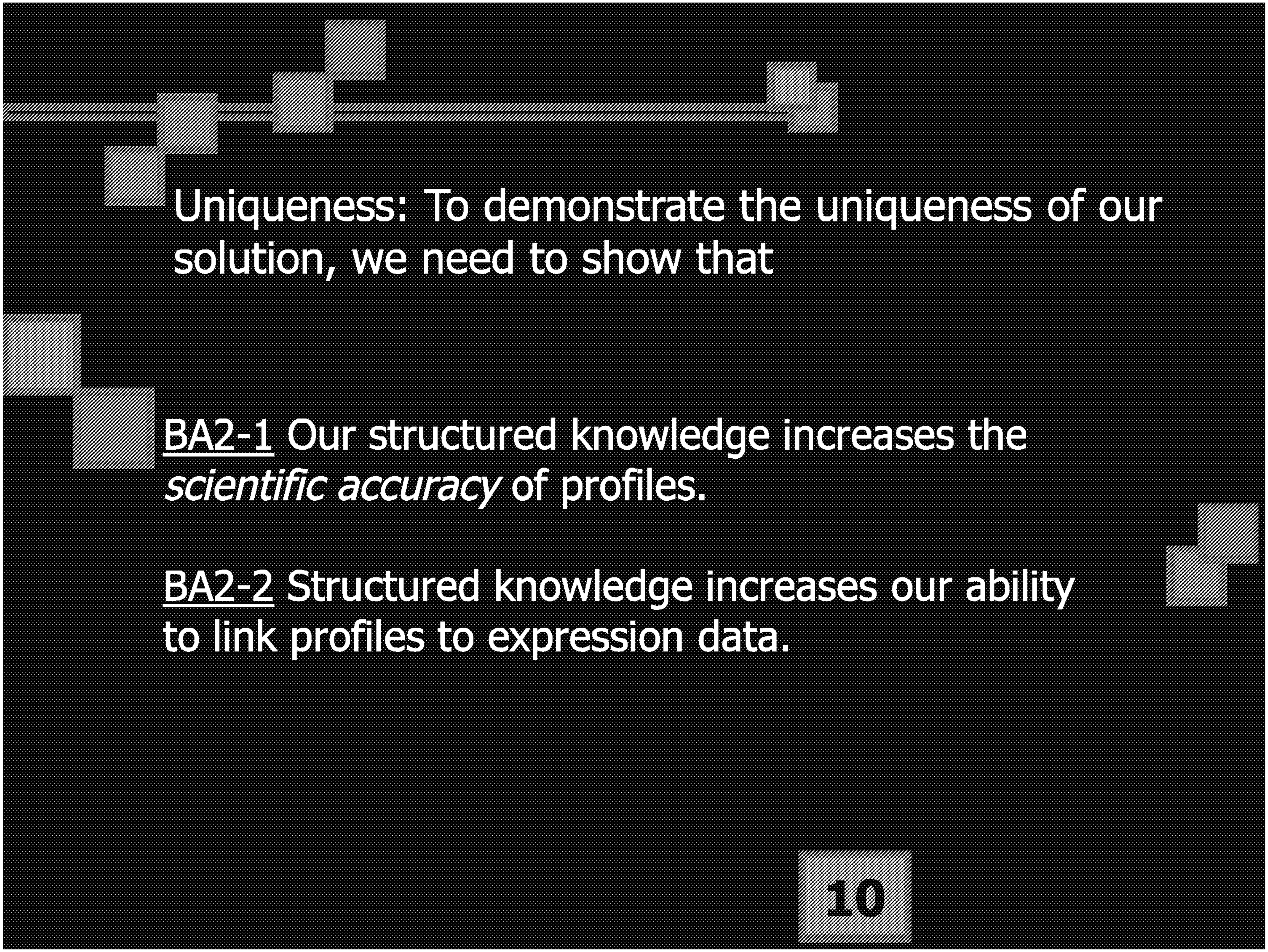
BA1-1: BCP/Profile modeling is biologically credible

BA1-2 Expression Analysis Results Using Profiles are biologically believable

BA1-3 Results are non-random

BA1-4 Results are novel

EXHIBIT 4 Con't

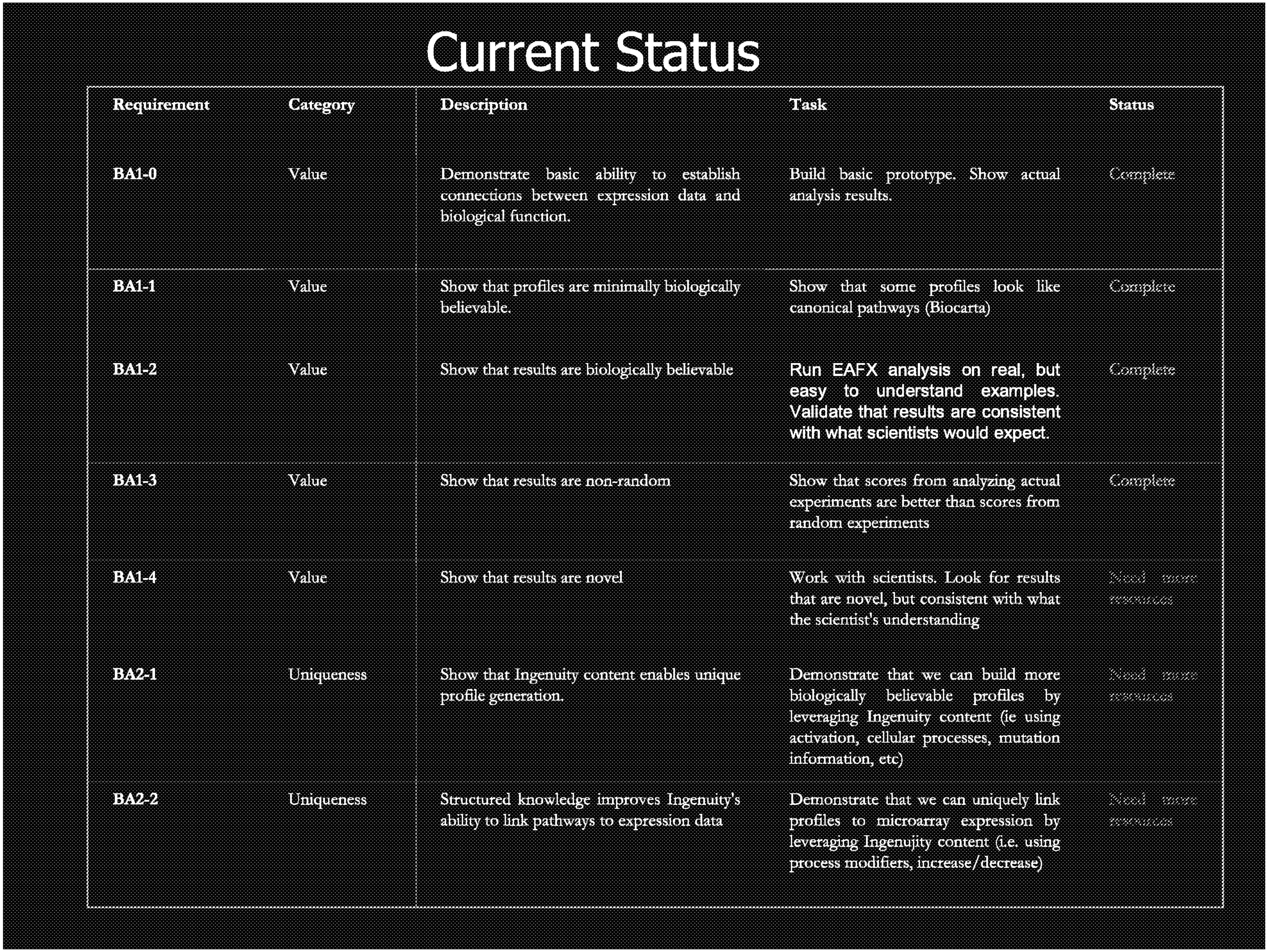


Uniqueness: To demonstrate the uniqueness of our solution, we need to show that

BA2-1 Our structured knowledge increases the *scientific accuracy* of profiles.

BA2-2 Structured knowledge increases our ability to link profiles to expression data.

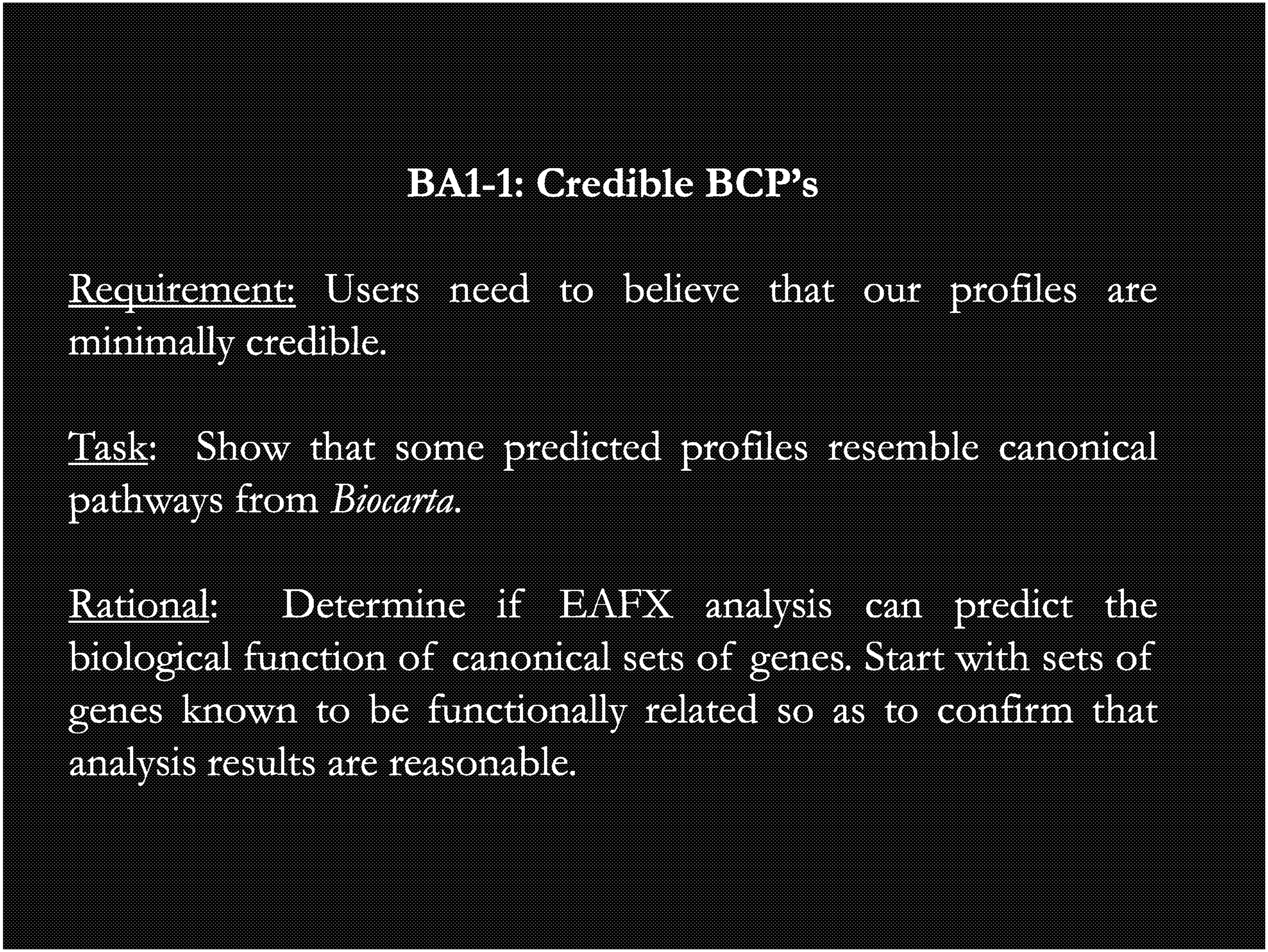
EXHIBIT 4 Con't



Current Status

Requirement	Category	Description	Task	Status
BA1-0	Value	Demonstrate basic ability to establish connections between expression data and biological function.	Build basic prototype. Show actual analysis results.	Complete
BA1-1	Value	Show that profiles are minimally biologically believable.	Show that some profiles look like canonical pathways (Biocarta)	Complete
BA1-2	Value	Show that results are biologically believable	Run EAFX analysis on real, but easy to understand examples. Validate that results are consistent with what scientists would expect.	Complete
BA1-3	Value	Show that results are non-random	Show that scores from analyzing actual experiments are better than scores from random experiments	Complete
BA1-4	Value	Show that results are novel	Work with scientists. Look for results that are novel, but consistent with what the scientist's understanding	Need more resources
BA2-1	Uniqueness	Show that Ingenuity content enables unique profile generation.	Demonstrate that we can build more biologically believable profiles by leveraging Ingenuity content (ie using activation, cellular processes, mutation information, etc)	Need more resources
BA2-2	Uniqueness	Structured knowledge improves Ingenuity's ability to link pathways to expression data	Demonstrate that we can uniquely link profiles to microarray expression by leveraging Ingenujity content (i.e. using process modifiers, increase/decrease)	Need more resources

EXHIBIT 4 Con't



BA1-1: Credible BCP's

Requirement: Users need to believe that our profiles are minimally credible.

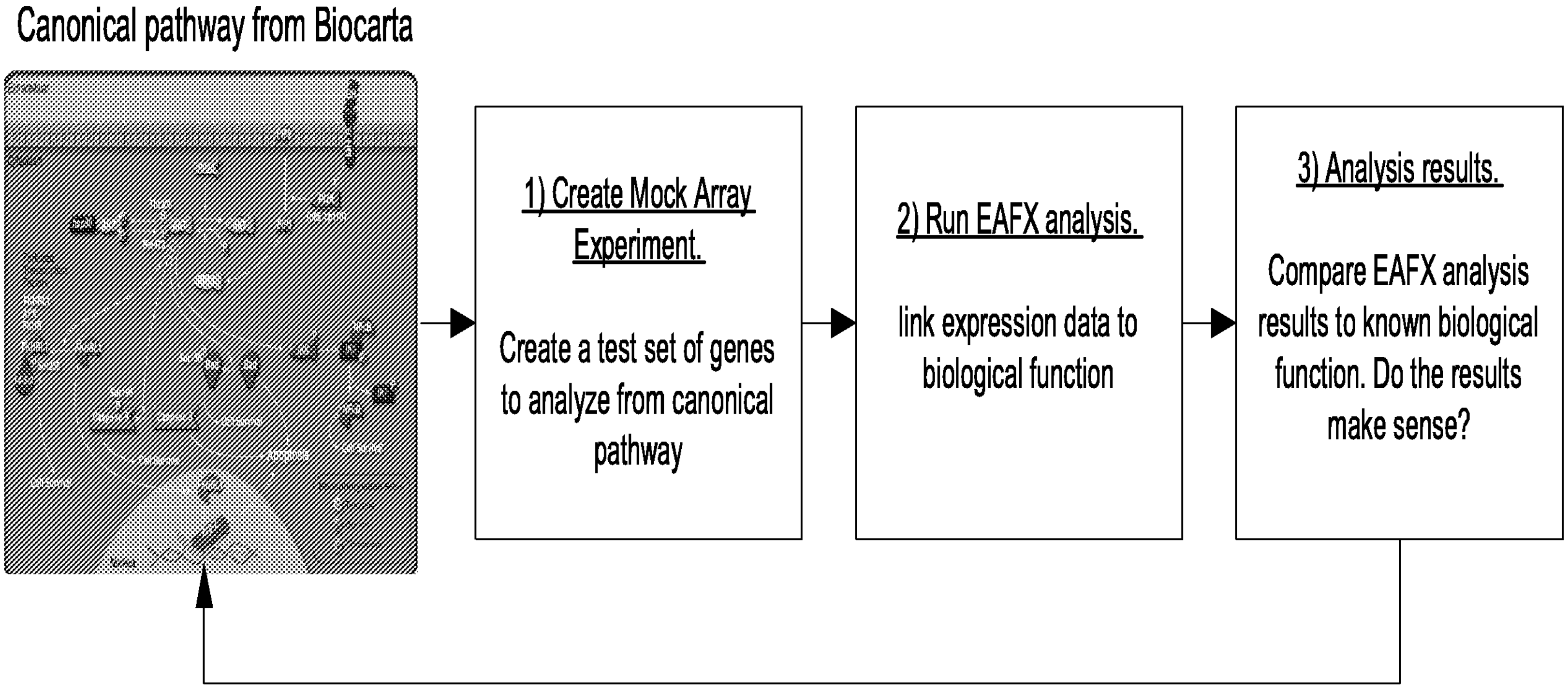
Task: Show that some predicted profiles resemble canonical pathways from *Biocarta*.

Rational: Determine if EAFX analysis can predict the biological function of canonical sets of genes. Start with sets of genes known to be functionally related so as to confirm that analysis results are reasonable.

EXHIBIT 4 Con't

BA1-1 -continued

Comparing EAFX to Canonical Pathways



Canonical pathway from Biocarta

1) Create Mock Array Experiment.

Create a test set of genes to analyze from canonical pathway

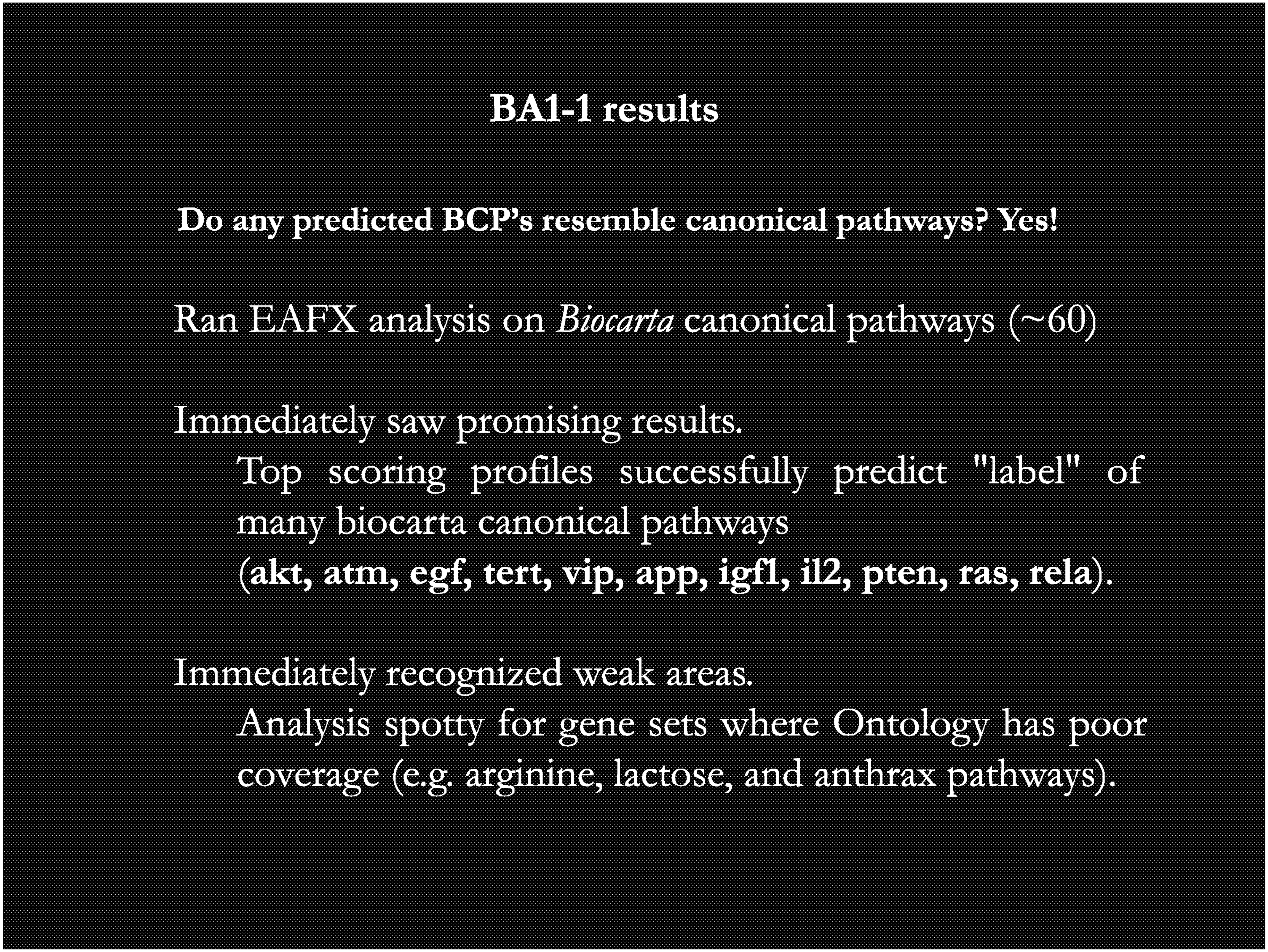
2) Run EAFX analysis.

link expression data to biological function

3) Analysis results.

Compare EAFX analysis results to known biological function. Do the results make sense?

EXHIBIT 4 Con't



BA1-1 results

Do any predicted BCP's resemble canonical pathways? Yes!

Ran EAFX analysis on *Biocarta* canonical pathways (~60)

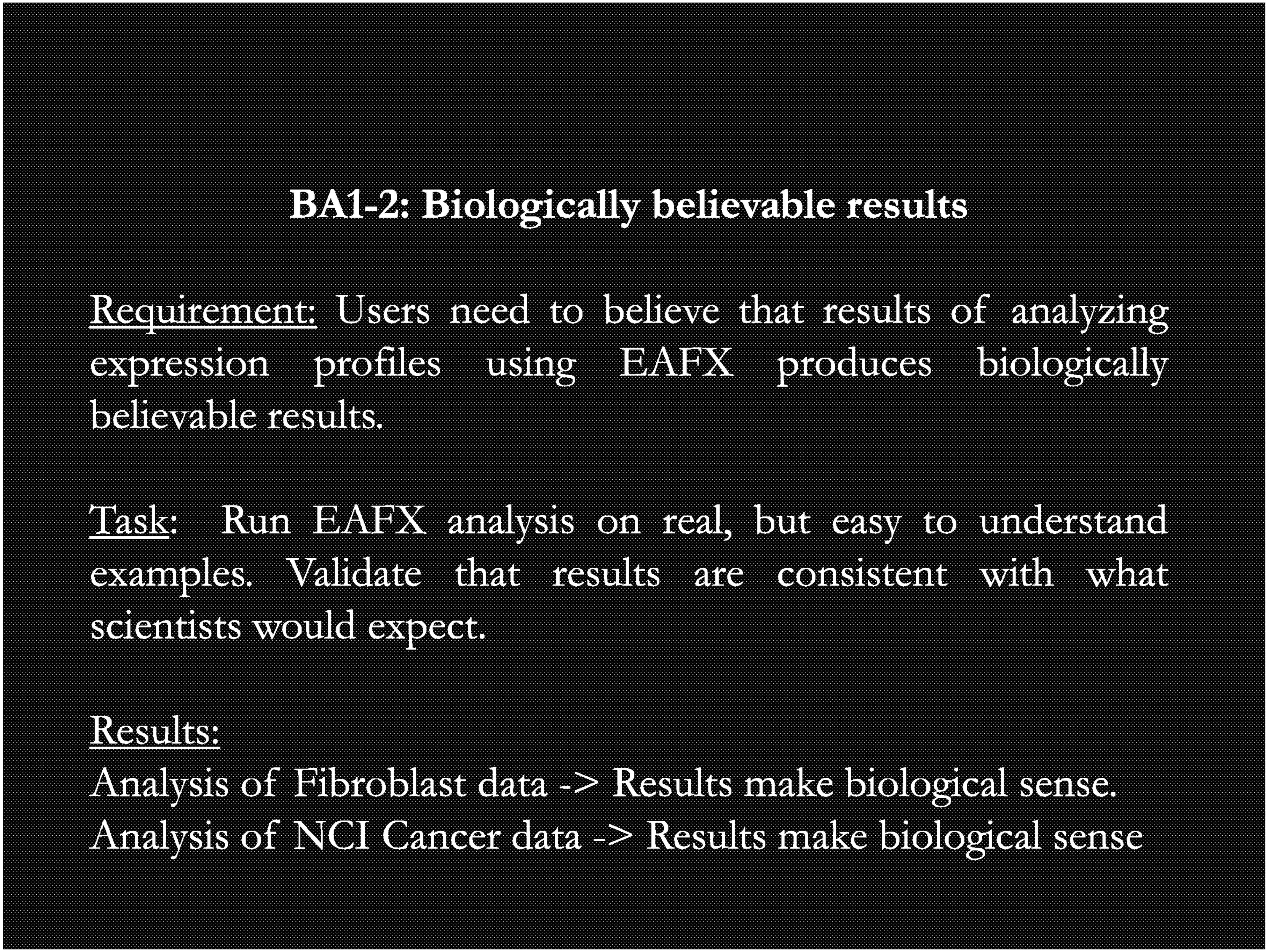
Immediately saw promising results.

Top scoring profiles successfully predict "label" of many biocarta canonical pathways
(**akt, atm, egf, tert, vip, app, igf1, il2, pten, ras, rela**).

Immediately recognized weak areas.

Analysis spotty for gene sets where Ontology has poor coverage (e.g. arginine, lactose, and anthrax pathways).

EXHIBIT 4 Con't



BA1-2: Biologically believable results

Requirement: Users need to believe that results of analyzing expression profiles using EAFX produces biologically believable results.

Task: Run EAFX analysis on real, but easy to understand examples. Validate that results are consistent with what scientists would expect.

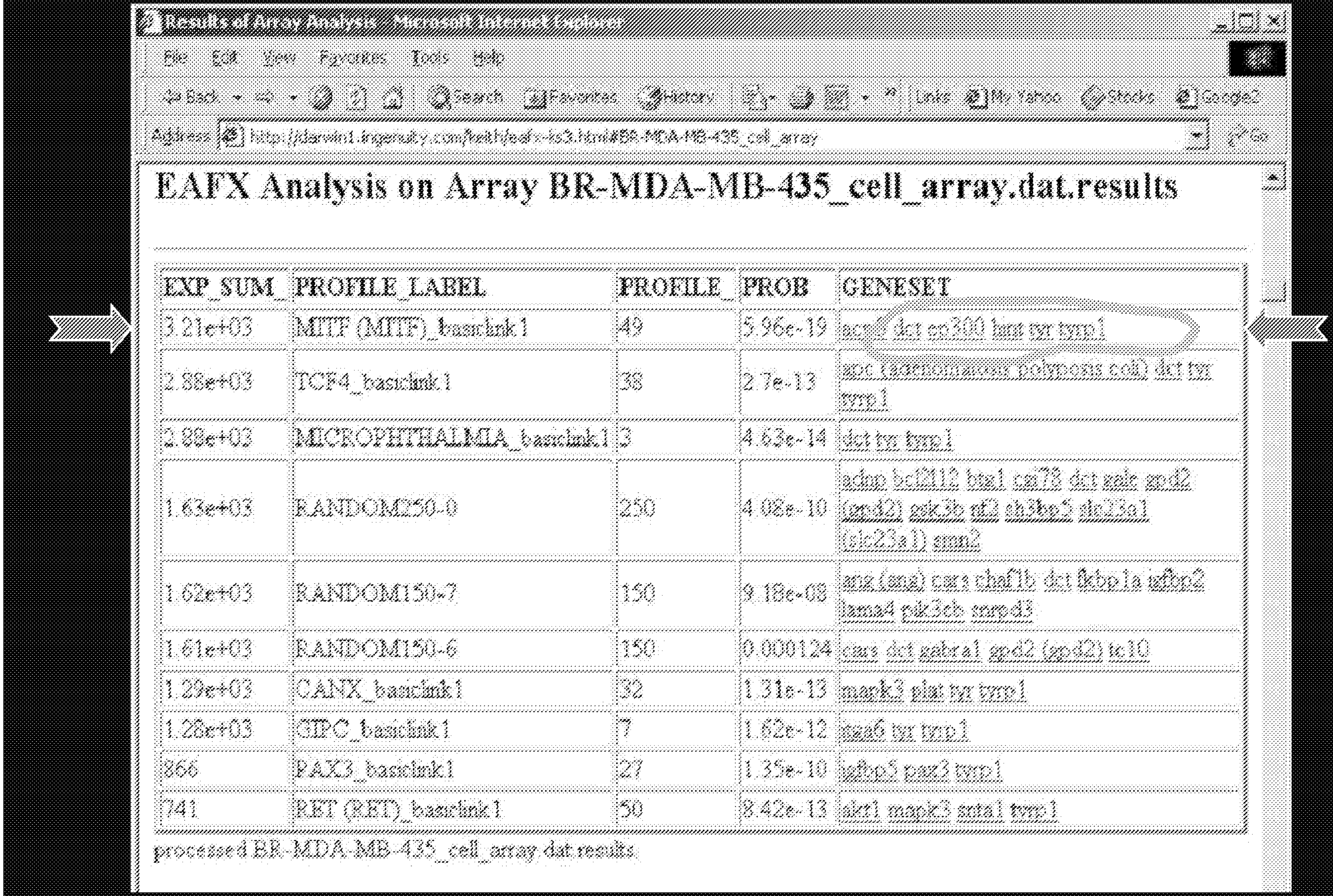
Results:

Analysis of Fibroblast data -> Results make biological sense.

Analysis of NCI Cancer data -> Results make biological sense

EXHIBIT 4 Con't

Gene-Centric Profile Scoring for Cancer Cell Line


Results of Array Analysis - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History Links My Yahoo Stocks Google2
Address http://darwin1.ingenuity.com/keith/eafx-ks3.html#BR-MDA-MB-435_cell_array
Go

EAFX Analysis on Array BR-MDA-MB-435_cell_array.dat.results

EXP_SUM_	PROFILE_LABEL	PROFILE_	PROB	GENESET
3.21e+03	MITF (MITF)_basiclink1	49	5.96e-19	acn5 dct ep300 hmt tyr tyrp1
2.88e+03	TCF4_basiclink1	38	2.7e-13	apc (adenomatous polyposis coli) dct tyr tyrp1
2.88e+03	MICROPHTHALMIA_basiclink1	3	4.63e-14	dct tyr tyrp1
1.63e+03	RANDOM250-0	250	4.08e-10	adnp bcl2l12 btg1 cai78 dct galc gpd2 (gpd2) gsk3b nf2 sh3bp5 slc23a1 (slc23a1) smn2
1.62e+03	RANDOM150-7	150	9.18e-08	anx (ana) cars chaf1b dct fkbp1a igfbp2 lama4 pik3cb snrpd3
1.61e+03	RANDOM150-6	150	0.000124	cars dct gabra1 gpd2 (gpd2) tcl10
1.29e+03	CANX_basiclink1	32	1.31e-13	mapk3 plat tyr tyrp1
1.28e+03	GIPC_basiclink1	7	1.62e-12	naa6 tyr tyrp1
866	PAX3_basiclink1	27	1.35e-10	igfbp5 pax3 tyrp1
741	RET (RET)_basiclink1	50	8.42e-13	akt1 mapk3 snta1 tyrp1

processed BR-MDA-MB-435_cell_array.dat.results

EXHIBIT 4 Con't

Gene-Centric Profile Scoring: Consistent with Biology

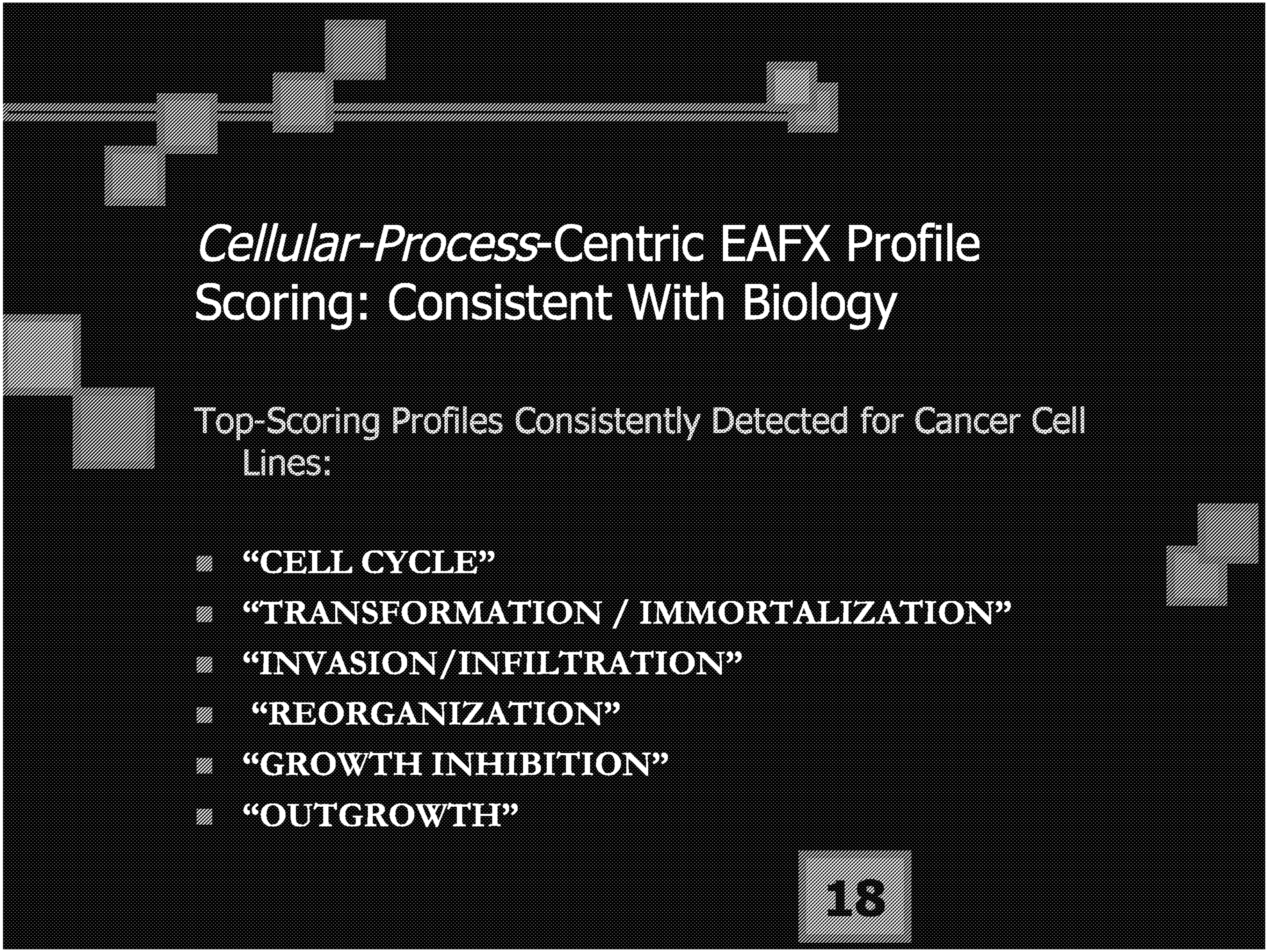
MITF Profile Members (select examples):

- Dct
- Ep300
- Hint
- Tyr
- Tyrp1

Their Ontological Assertions Relevant to Cancer:

- **GROWTH OF TUMORS**
- **EXPRESSION IN TUMORS**
- **P53: BINDING, RESPONSE STABILIZATION, ACCUMULATION**
- **SURVIVAL OF TUMOR CELLS**
- **CELL CYCLE**
- **GROWTH/PROLIFERATION OF TUMORS & CELL LINES**
- **TRANSFORMATION / IMMORTALIZATION**
- **EXPRESSION IN OTHER TRANSFORMED CELL LINES (e.g. HeLA, Melanoma)**

EXHIBIT 4 Con't

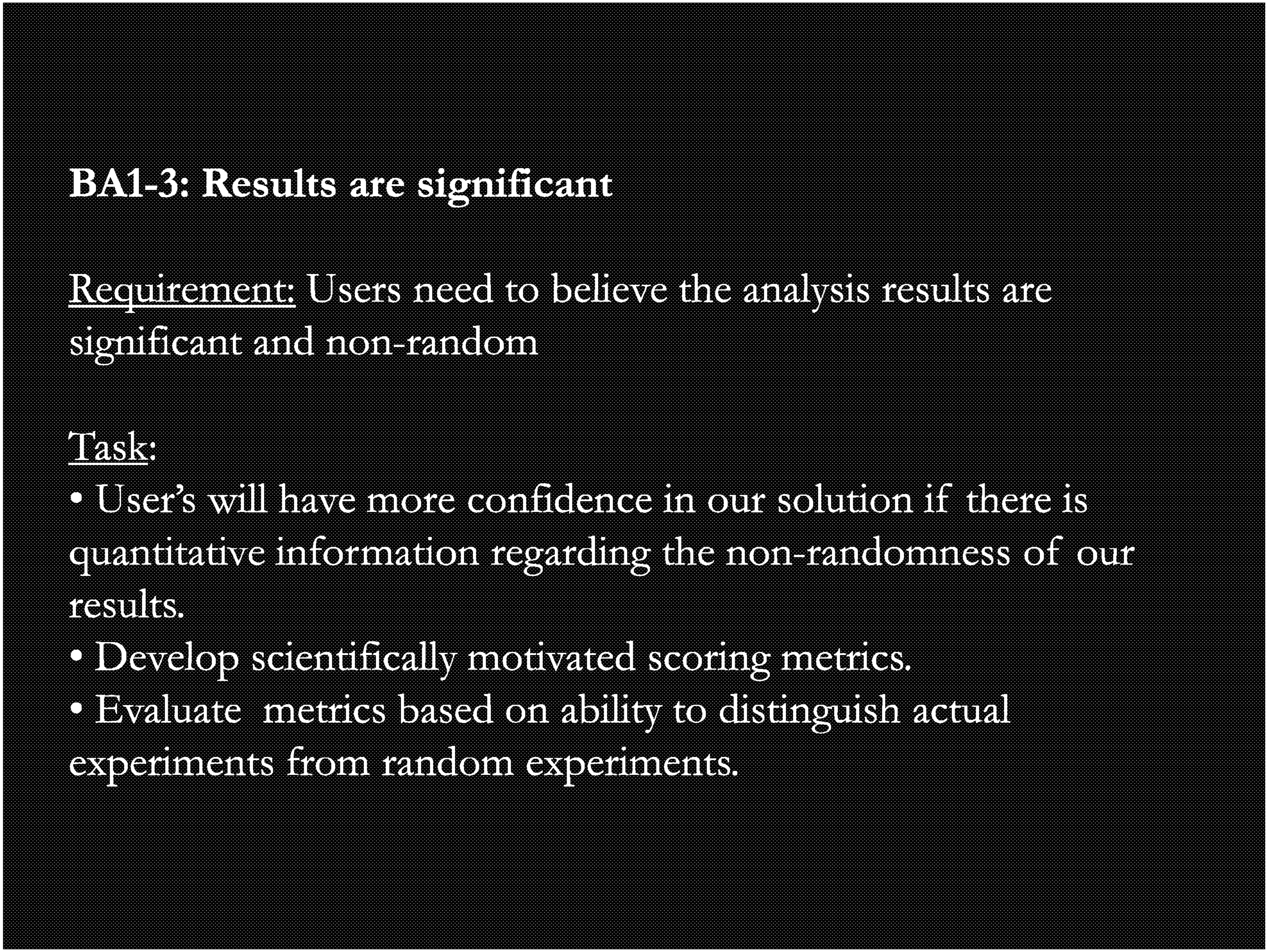


Cellular-Process-Centric EAFX Profile Scoring: Consistent With Biology

Top-Scoring Profiles Consistently Detected for Cancer Cell Lines:

- **"CELL CYCLE"**
- **"TRANSFORMATION / IMMORTALIZATION"**
- **"INVASION/INFILTRATION"**
- **"REORGANIZATION"**
- **"GROWTH INHIBITION"**
- **"OUTGROWTH"**

EXHIBIT 4 Con't



BA1-3: Results are significant

Requirement: Users need to believe the analysis results are significant and non-random

Task:
- User’s will have more confidence in our solution if there is quantitative information regarding the non-randomness of our results.
- Develop scientifically motivated scoring metrics.
- Evaluate metrics based on ability to distinguish actual experiments from random experiments.

EXHIBIT 4 Con't

BA1-3: Results are non-random

Stanford Fibroblast Microarray data.

- Downloaded array data from 12 time points (15 min. - 24 hr).
- Generated 100 shuffled/randomized variants of each array experiment.
- Matches between Ingenuity profiles and expression data that seem to be non-random are found.
- The degree of non-randomness appears to be much stronger in real data.

Analysis scores from real data in red. Analysis scores from randomized data in black.

Top 10 out of 100 sorted results shown below.

Array Exp	.25h	.5h	1h	2h	4h	6h	8h	12h	16h	20h	24h
Score	40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	101	100	113
	[illegible]	52	64	55	96	92	110	96	95	90	103
	24	48	59	55	92	86	101	96	[illegible]	86	96
	24	44	57	53	90	86	101	92	86	84	93
	23	43	56	50	89	83	98	91	85	83	88
	23	43	55	48	87	79	98	89	85	82	86
	22	40	51	47	84	79	96	87	82	82	84
	22	38	51	46	83	79	96	86	81	82	82
	22	37	50	46	82	79	91	85	79	82	81

EXHIBIT 4 Con't

BA1-3: Results are non-random

<u>NCI Cancer data.</u> Use data that is more realistic. Early results were non-ideal. We worked on more sophisticated scoring functions and improved our ability to discriminate real from random data.

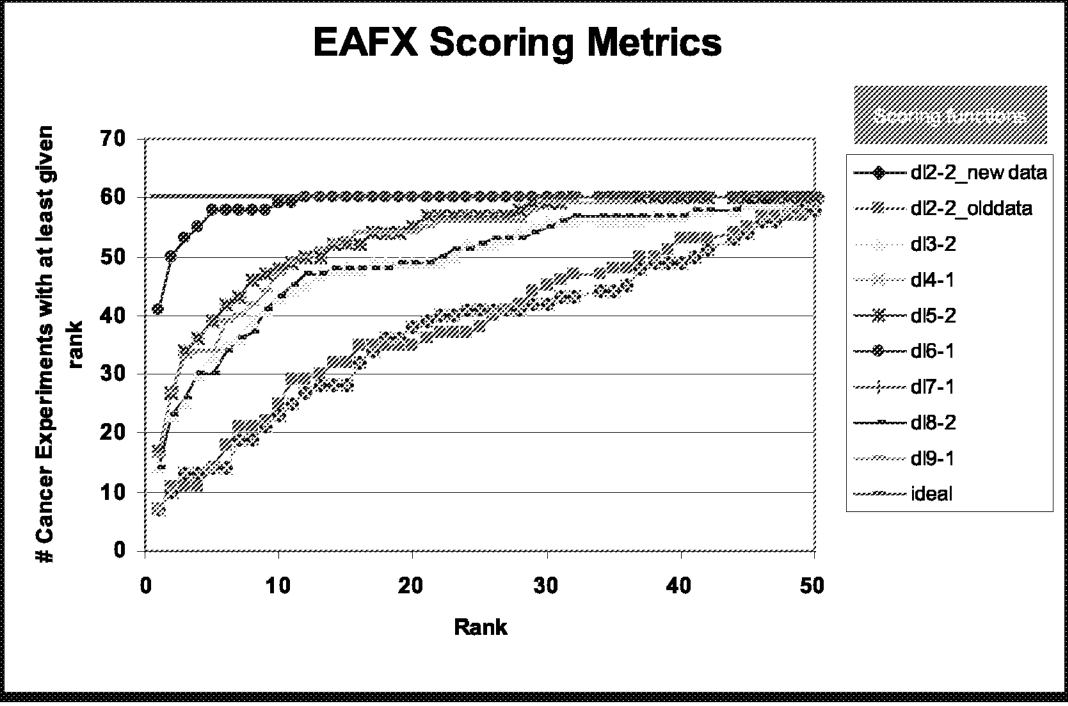
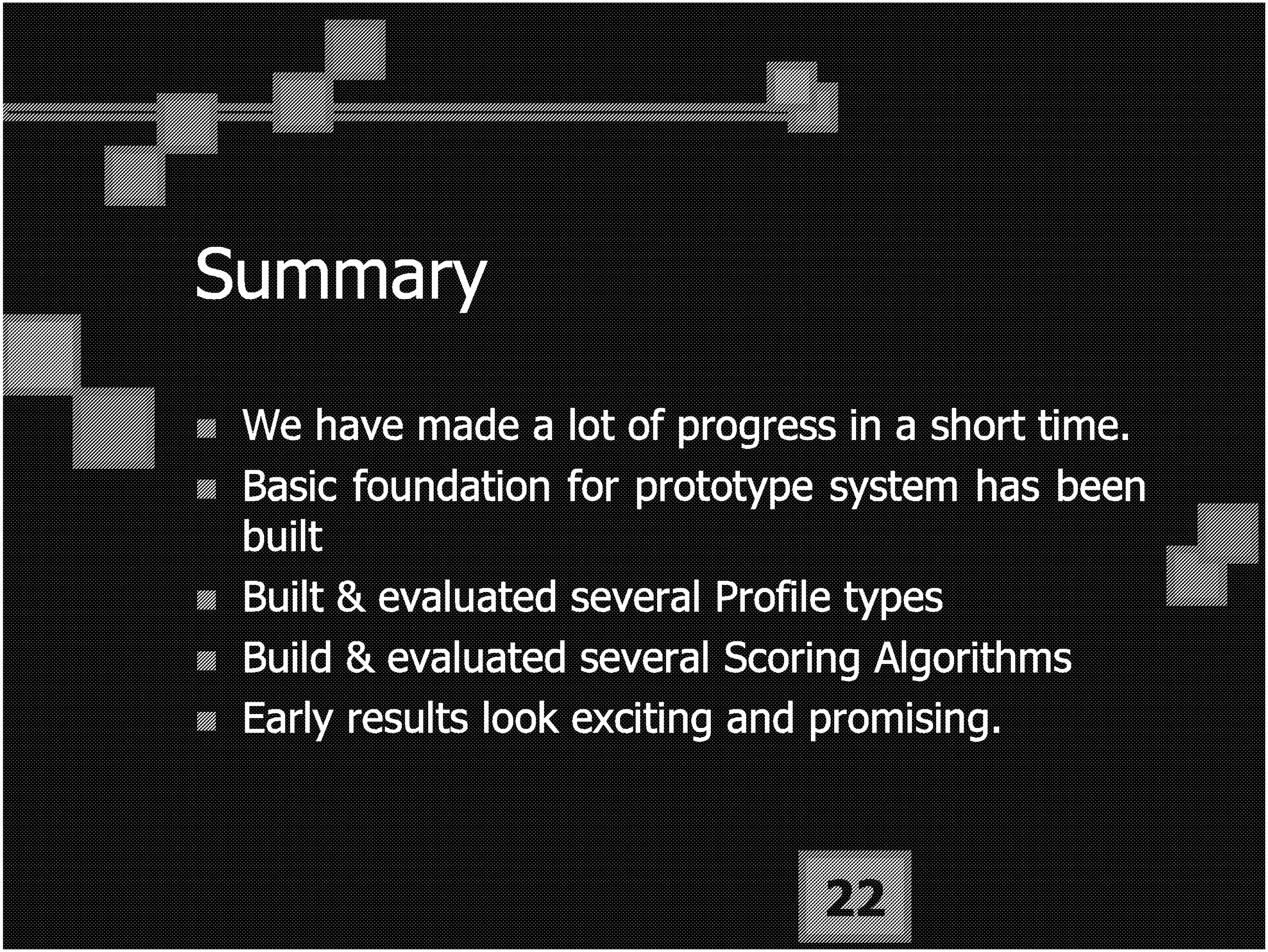


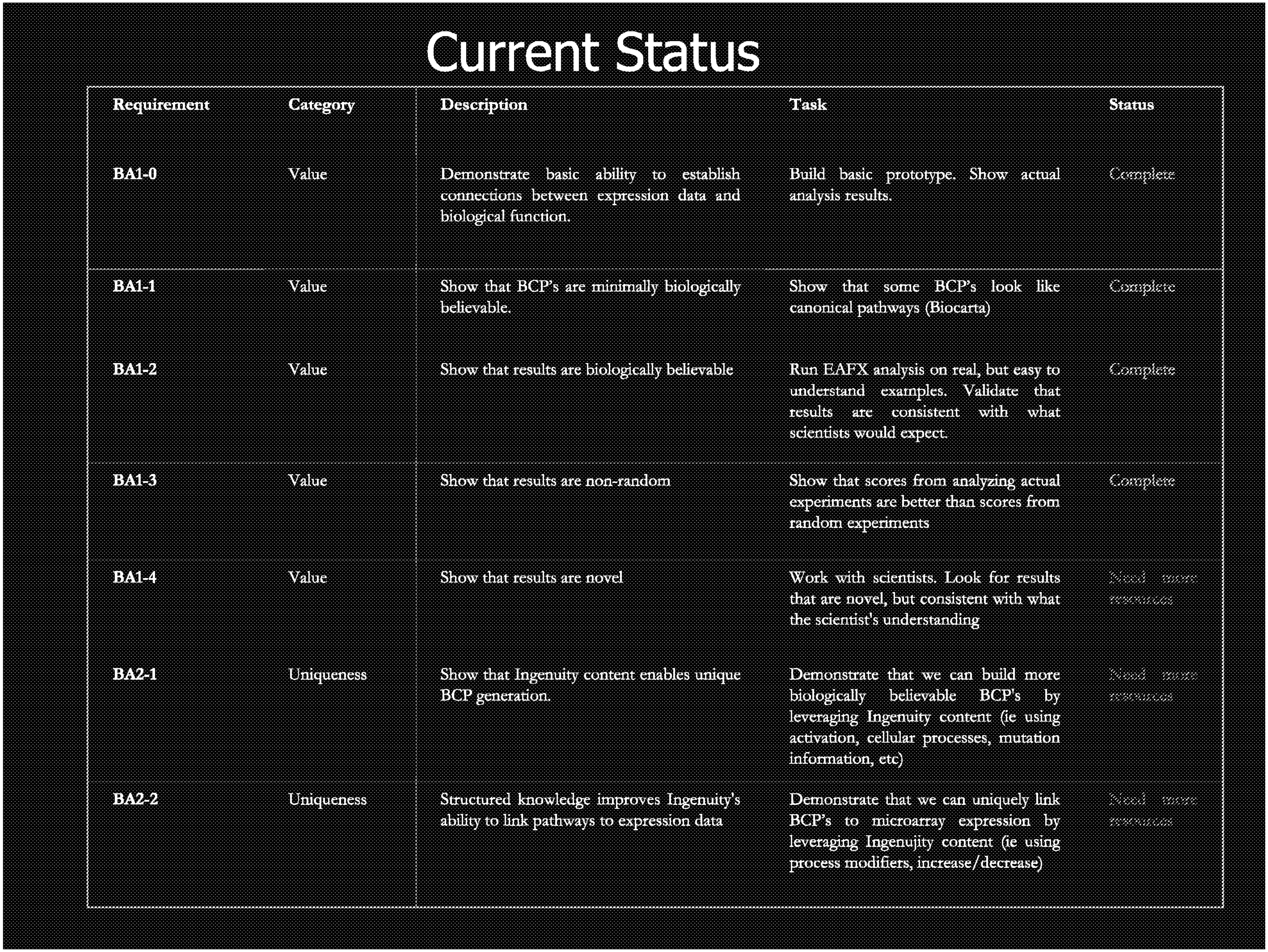
EXHIBIT 4 Con't



Summary

- We have made a lot of progress in a short time.
- Basic foundation for prototype system has been built
- Built & evaluated several Profile types
- Build & evaluated several Scoring Algorithms
- Early results look exciting and promising.

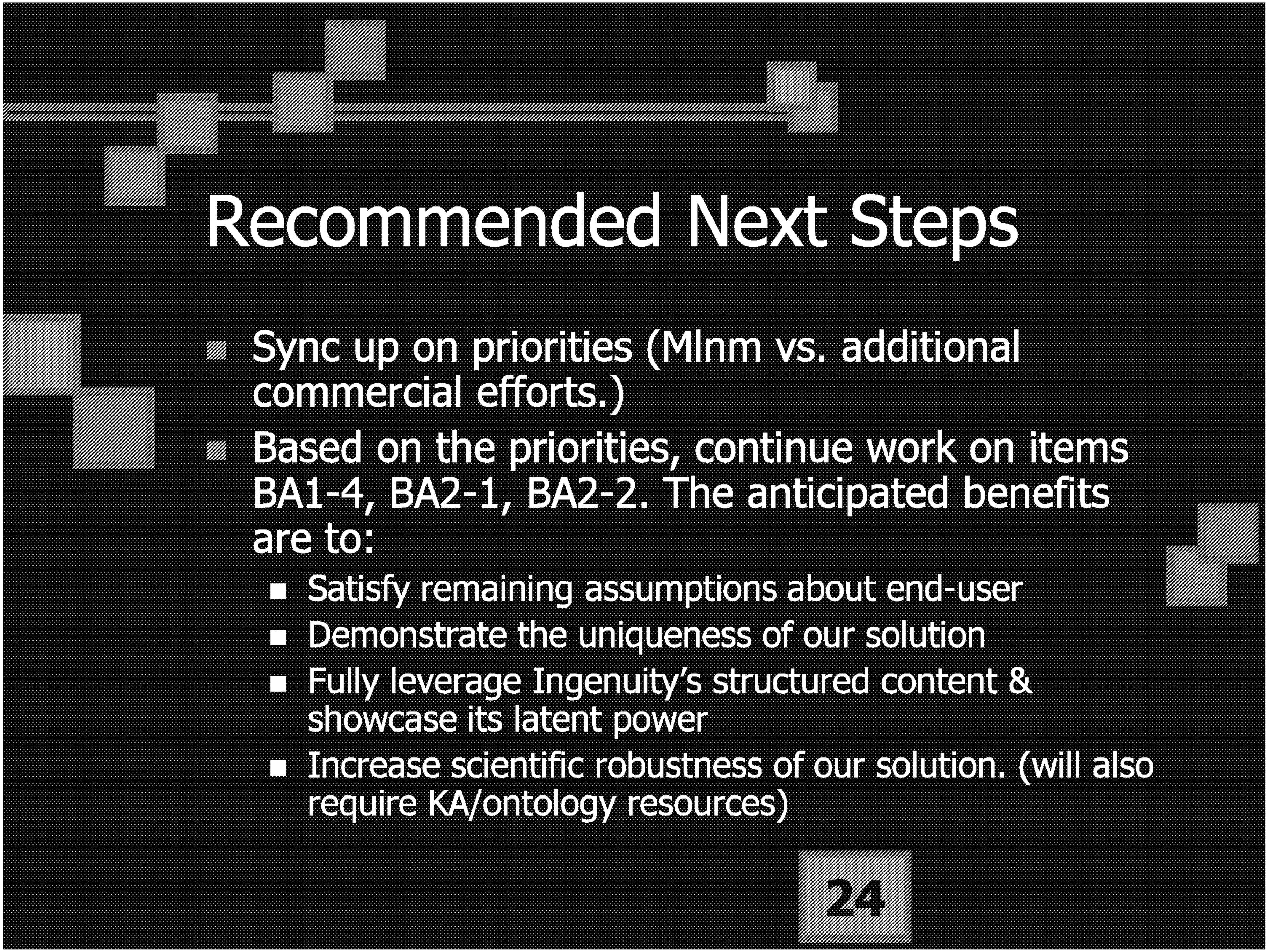
EXHIBIT 4 Con't



Current Status

Requirement	Category	Description	Task	Status
BA1-0	Value	Demonstrate basic ability to establish connections between expression data and biological function.	Build basic prototype. Show actual analysis results.	Complete
BA1-1	Value	Show that BCP's are minimally biologically believable.	Show that some BCP's look like canonical pathways (Biocarta)	Complete
BA1-2	Value	Show that results are biologically believable	Run EAFX analysis on real, but easy to understand examples. Validate that results are consistent with what scientists would expect.	Complete
BA1-3	Value	Show that results are non-random	Show that scores from analyzing actual experiments are better than scores from random experiments	Complete
BA1-4	Value	Show that results are novel	Work with scientists. Look for results that are novel, but consistent with what the scientist's understanding	Need more resources
BA2-1	Uniqueness	Show that Ingenuity content enables unique BCP generation.	Demonstrate that we can build more biologically believable BCP's by leveraging Ingenuity content (ie using activation, cellular processes, mutation information, etc)	Need more resources
BA2-2	Uniqueness	Structured knowledge improves Ingenuity's ability to link pathways to expression data	Demonstrate that we can uniquely link BCP's to microarray expression by leveraging Ingenujity content (ie using process modifiers, increase/decrease)	Need more resources

EXHIBIT 4 Con't



Recommended Next Steps

- Sync up on priorities (Mlnm vs. additional commercial efforts.)
- Based on the priorities, continue work on items BA1-4, BA2-1, BA2-2. The anticipated benefits are to:
  - Satisfy remaining assumptions about end-user
  - Demonstrate the uniqueness of our solution
  - Fully leverage Ingenuity's structured content & showcase its latent power
  - Increase scientific robustness of our solution. (will also require KA/ontology resources)